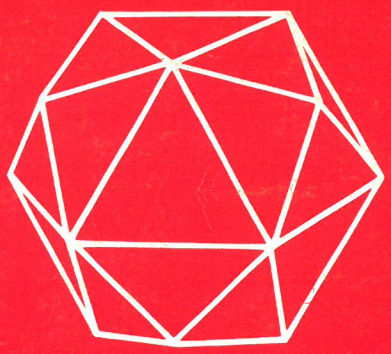
Quaderni di informatica

Rassegna di tecnologia ed applicazioni degli elaboratori elettronici
Anno XIV - Numero 3 - 1988

34



CENTRO STUDI
INFORMATICA E AUTOMAZIONE

ARCHIVIO STORICO
ASSOCIAZIONE POZZO DI MIELE
FPDM-41
RQDI-132

Honeywell Bull

FPDM-41

Honeywell Bull

Quaderni di informatica

Rassegna di tecnologia ed applicazioni degli elaboratori elettronici
Anno XIV - Numero 3 - 1988

Sommario

Informatica e innovazione: il caso del settore tessile-abbigliamento

L'informatica costituisce un fondamentale fattore di rivitalizzazione di settori cosidetti "maturi". Un esempio importante è quello del tessile italiano, documentato in questa ricerca.
R. CAMAGNI, R. RABELLOTTI

Le database machine

L'impiego di sistemi espressamente dedicati alla gestione di basi di dati viene proposto da tempo. L'articolo illustra le possibili realizzazioni ed i pro e contro di questa soluzione.
L. FELICIAN, A. PAIUSCO

Il software didattico nella formazione scolastica e professionale

La didattica col computer viene discussa in questo articolo utilizzando diversi schemi di riferimento. Ne esce una analisi penetrante che illumina il soggetto di nuova luce.
M. PELLEREY

Sistema esperto per il progetto di reti di comunicazione

Il dimensionamento di una rete di trasmissione dati è un problema molto complesso. Per la sua soluzione si può oggi ricorrere alle tecniche della intelligenza artificiale.
M.B. GOZZOLI, G. ZANOTTI

Interazione vocale col computer

Colloquiare a voce con la macchina presenta oggi ancora molte limitazioni. Il progresso in questo settore passa, in particolare, attraverso una migliore comprensione dei processi di comunicazione umana.
J. WATERWORTH

Pensare in parallelo

L'organizzazione seriale che caratterizza gli attuali computer riflette, secondo l'Autore, schemi culturali tipici dell'occidente. Ma il trend verso soluzioni parallele può prendere spunti dalle culture orientali.
T. HUETER

Direttore Responsabile
Franco Filippazzi

Comitato di Redazione
A. Cicu, C. Falcetti, P. Lupo,
M. Nobile, G. Occhini
F. Sala

Redazione
Honeywell Bull
Centro Studi Informatica e Automazione
Via G.M. Vida, 11 - 20127 Milano

Stampa
tipolito Maggioni s.a.s.

Autorizzazione del Tribunale di Milano
n. 93 del 20 Marzo 1974

«Quaderni di Informatica» ospita articoli e contributi di varia provenienza. La responsabilità delle opinioni espresse rimane agli Autori.

La riproduzione di articoli della rivista, o di parte di essi, è consentita solo citando la fonte e l'autore.

Informatica e innovazione: il caso del settore tessile-abbigliamento

ROBERTO CAMAGNI,
ROBERTA RABELLOTTI (*)

Università Bocconi
Milano

1. L'ambito d'indagine

L'opinione diffusa fino ad alcuni anni fa che le produzioni "mature" fossero destinate ad essere trasferite nei paesi meno industrializzati, si fonda essenzialmente sull'ampia disponibilità in tali paesi di manodopera a basso costo per lavori ripetitivi, come quelli richiesti da una produzione fortemente standardizzata. Questa interpretazione del modello di sviluppo, legata alle teorie tradizionali della divisione internazionale del lavoro e del ciclo di vita del prodotto, non sembra più essere l'unica in grado di spiegare la situazione economica mondiale, nè tanto meno quella italiana.

In particolare, le produzioni tessili rappresentano un terreno cruciale di definizione dei rapporti tra paesi industrializzati e paesi in via di sviluppo. Il problema della ristrutturazione industriale e del "redeployment" si pongono in questo settore con particolare drammaticità. A questo riguardo sorge il problema di capire fino a che punto queste produzioni possano considerarsi tecnologicamente "mature" e quale possa essere il ruolo delle nuove tecnologie micro-elettroniche nel definire le prospettive di sviluppo del settore.

In Italia, il tessile/abbigliamento si sviluppa negli anni '50 basandosi su solide capacità artigianali, su una tradizione di grande creatività e su un vasto mercato del lavoro femminile da cui attingere manodopera a basso costo.

Negli anni '60 consulenti stranieri introducono criteri di industrializzazione che trasformano i laboratori in industrie. In questo periodo nascono i grandi colossi dell'abbigliamento, come Gft, Marzotto, Lebole, che si caratterizzano per prodotti di tipo classico, standardizzato, con poche varietà di modello e tessuto. Agli inizi degli anni '70 si sviluppano forme di decentramento e il settore si struttura su alcune grandi imprese e un elevatissimo numero di piccole e medie imprese con accentuata "nati-mortalità", consentita dalla mancanza di barriere all'entrata tecnico-organizzative e finanziarie.

Negli ultimi anni, la strategia imprenditoriale italiana di "adattamento/resistenza" alla penetrazione dei paesi in via di sviluppo in campo tessile/abbigliamento si è incentrata sull'innovazione tecnologica e organizzativa e sulla estensione del ciclo produttivo alle fasi più vicine al prodotto finale e quindi a più alto valore aggiunto.

Nell'abbigliamento si è puntato soprattutto sull'innovazione di prodotto, sulla sua qualità e immagine e sulle politiche di commercializzazione. La capacità tecnologica e la conservazione di tutte le fasi del ciclo, anche di quelle meno nobili, unite alla creatività e alla imprenditorialità di alcune realtà locali hanno costituito la base di partenza per il successo dell'industria tessile italiana.

Qui di seguito si espongono i risultati dello studio di una serie di casi di adozione di successo di innovazioni tecnologiche e organizzative. A questo scopo sono state individuate 8 imprese [1], tra le più avanzate tecnologicamente all'interno del proprio comparto produttivo, al fine di analizzare l'adozione di innovazioni nei processi produttivi e nei prodotti e l'impatto di questa sulla struttura macro-organizzativa.

L'obiettivo, attraverso l'analisi di alcuni esempi di "best practice" tecnologica e organizzativa, probabilmente destinata a divenire nel corso di un certo numero di anni "average practice", è quello di portare alcune riflessioni e, per quanto possibile, delle previsioni in merito alla crescita della competitività delle imprese del settore, alla possibilità di innovazione nelle forme organizzative e all'impiego di nuove tecnologie come stimolo e occasione di innovazioni di prodotto.

2. L'introduzione di tecnologie informatiche

Le tecnologie a base elettronica si adattano molto bene alla caratteristica variabilità e differenziazione

(*) Questo articolo costituisce la sintesi di una ricerca svolta presso l'Istituto di Economia della Università Bocconi di Milano per Honeywell Bull Italia, diretta da R. Camagni e svolta da R. Rabellotti. I paragrafi 1 e 2 sono stati redatti da R. Rabellotti e il paragrafo 3 congiuntamente dai due autori.

(1) Le imprese intervistate sono: Gruppo Finanziario Tessile, Gruppo Tessile Miroglio, Loropiana, Max Mara, Missoni, Nanibon, Ratti, Stamperia Di Camerlate.

interna del settore tessile/abbigliamento. Le esigenze tecnologiche sono infatti molto diverse a seconda della dimensione d'azienda e delle possibilità finanziarie, della ampiezza dei campionari, del tipo di prodotto, delle tendenze della moda, del tipo di mercato e dei metodi di commercializzazione [12].

2.1 L'innovazione nei processi produttivi

L'introduzione delle tecnologie elettroniche all'interno dei processi produttivi nel settore tessile/abbigliamento interessa, in modo più o meno consistente, ogni fase del ciclo. Come si può osservare dall'analisi della figura 1 tutte le fasi di cui ci occupiamo - la tessitura, la stampa, la tintura, il finissaggio e la confezione - sono soggette ad importanti fenomeni di rinnovamento attraverso l'introduzione di tecnologie in parte consolidate, in parte in fase di industrializzazione ed in parte solamente in fase di studio.

Tra le tecnologie più consolidate rientrano sicuramente i sistemi CAD (Computer Aided Design) che sono stati infatti adottati da tutte le imprese del nostro campione.

Nell'ambito del settore tessile il CAD può avere applicazioni diverse: può essere infatti utilizzato per la progettazione del tessuto, per lo sviluppo dei disegni da stampare, per la ricettazione dei colori ed infine nella confezione per la progettazione dei modelli, lo sviluppo delle taglie e l'organizzazione della fase di taglio [11].

I vantaggi che derivano dall'adozione di sistemi CAD sono numerosi: il primo effetto riguarda la fase puramente creativa del ciclo produttivo. Gli stilisti, per mezzo del CAD, possono infatti disegnare numerose varianti dello stesso modello, correggerle e analizzarle in tempi molto brevi e senza costi di materie prime.

Un esperimento molto interessante a questo riguardo si sta realizzando presso il CITER (Centro Informazione Tessile dell'Emilia Romagna) di Carpi, in collaborazione con l'ENEA.

L'obiettivo è la creazione di una stazione di lavoro basata su personal computer e videodisco, dove accanto ad un sistema CAD, si sta creando una banca dati di immagini per gli operatori del tessile/abbigliamento. L'ENEA ha infatti elaborato un progetto per l'archiviazione su videodisco di immagini attinenti al settore secondo vari criteri di classificazione: l'anno e la stagione di riferimento, la tendenza, la linea, i colori, la forma dei dettagli, etc.

L'archivio che si sta costituendo rappresenta la prima banca dati di immagini per applicazioni industriali prodotta in Italia ed è costituita da immagini fotografiche, disegni ed informazioni trascritte, sui quali sarà possibile lavorare mediante una tavoletta grafica. Le caratteristiche del data base - "user-friendliness", "off-the-shelf technology" e prezzo di consultazione non elevato - dovrebbero permetterne un uso su larga scala, anche in un contesto produttivo caratterizzato da imprese di piccola e media dimensione, quale quello emiliano-romagnolo.

Un'altra importante conseguenza dell'adozione del CAD è il risparmio di materie prime e tempi di lavorazione nella fase di produzione dei campionari. Tale risparmio è certamente sensibile, rapportato al fatto che la realizzazione del campionario rappresenta una voce importante nel bilancio di spesa delle imprese tessili.

Tuttavia, più ancora che per il risparmio in termini economici, l'accorciamento dei tempi di produzione dei campionari ha un valore strategico molto elevato, poichè permette di abbreviare i tempi che intercorrono tra lo sviluppo delle nuove idee-moda e i termini di presentazione delle collezioni.

L'obiettivo finale delle imprese sarebbe quello di eliminare completamente il campionario tradizionale, sostituendolo con una serie di immagini su calcolatore.

Tuttavia difficoltà di risoluzione grafica impediscono in molti casi la sostituzione e rendono necessaria la produzione del campionario finale da mostrare al cliente.

Infine i sistemi CAD permettono di fornire un migliore servizio commerciale, di offrire cioè prodotti maggiormente "custom-oriented" poichè il cliente può vedere, giudicare ed intervenire su di essi ancora prima che vengano realizzati.

L'integrazione del CAD con sistemi CAM (Computer Aided Manufacturing) rappresenta un ulteriore passo verso l'automazione nel settore; essa può avvenire mediante interfacciamento con i telai, per la produzione automatica dei campioni (una impresa del campione lo ha in fase di studio); con lo spettrofotometro per la misurazione dei colori e con "cucine colori" automatiche (adottate da 3 imprese) ed infine con macchine per il taglio (2 imprese) e il cucito automatico (3 imprese).

Di particolare interesse anche per i grandi sviluppi che si intravvedono, appaiono le applicazioni di sistemi CAD/CAM nelle fasi di stampa e tintoria. Per la stampa, in particolare nel comparto della seta, lo sviluppo dei lucidi comporta delle forti riduzioni di costi e di tempi di lavorazione. Nel campo della tintoria è di particolare interesse la "cucina colori" automatica che consiste nel laboratorio di preparazione e controllo dei colori, operazioni tradizionalmente svolte da abilissimi operai che miscelano manualmente polveri, coloranti liquidi e diluenti.

I nuovi sistemi completamente automatici garantiscono precisione e ripetibilità molto maggiori dei vecchi metodi artigianali, misurando e memorizzando ricette e dati sul consumo dei prodotti, misurando i colori per mezzo dello spettrofotometro e fornendo infine programmi per il calcolo di economicità delle differenti miscele per ottenere una data tinta.

Molto interessanti sono anche le applicazioni di sistemi CAD/CAM nella confezione, per esempio l'introduzione di macchine per il taglio con lame di un materasso (più teli sovrapposti, 10-15), per il taglio laser o ad acqua del telo singolo a secondo del tipo di tessuto utilizzato o l'introduzione di macchine a ciclo automatico per la cucitura, spesso dotate di programmatori elettronici.

Fig. 1 - La griglia dell'automazione tessile (esempi di tecnologie emergenti nell'area del tessile-abbigliamento).

	CAD	CAD-CAM	Informatica di processo	Robotica	Magazzini automatici
Tessitura	Progettazione del tessuto	• Creazione automatica dei campioni • Sviluppo delle matrici a cartone	Controllo integrato di qualità sui telai		
Stampa	• Sviluppo dei disegni • Analisi dei colori	Sviluppo dei lucidi di stampa	Controllo dei processi di lavaggio e stampa – a mano – a tavolo		
Tintura e finissaggio	Ricettazione e certificazione automatica dei colori su terminali	• Analisi dei colori con spettrofotometro e controllo dei colori "on line"	Controllo caldaie e processi di tintura		• Sistemi di pesatura, etichettatura e immagazzinaggio • Robot di carico e scarico delle macchine • Sistemi di movimentazione
Confezione – progettazione	• Progettazione modelli • Sviluppo delle taglie e piazzamento			Robot "Sarto" (studio Miti)	
– taglio	• Ottimizzazione del taglio	Taglio automatico – a lame – a laser – a acqua			
– cucito e stiro		Macchine automatiche	Controllo avanzamento lavori		

I vantaggi che derivano da tali adozioni sono la diminuzione dei tempi per prodotto, l'ottimizzazione dell'uso delle pezze riducendo al minimo gli scarti ed una maggiore omogeneità di prodotto [21].

Un ulteriore passo avanti verso l'automazione è costituito dall'*informatica di processo,* dal controllo cioè dell'intero ciclo produttivo, dalla fase di tessitura a quella di confezione. La messa a punto di un sistema di controllo in continuo è infatti ritenuta da tutte le imprese del campione fondamentale per aumentare l'affidabilità e la produttività dell'intero impianto e consentire la riduzione di manodopera per il rilevamento dei dati nei reparti di produzione, la riduzione delle fermate delle macchine, il miglioramento della qualità del prodotto, l'ottimizzazione della manutenzione preventiva e la conoscenza dei dati statistici e produttivi delle macchine, che si traduce in un miglioramento della gestione contabile e di magazzino. La conoscenza "in tempo reale" dello stato di avanzamento dei lavori permette inoltre un notevole incremento della flessibilità degli impianti, poichè rende possibili eventuali mutamenti delle caratteristiche del ciclo lungo il percorso compiuto dal prodotto, in funzione delle necessità di mercato.

L'applicazione dei *robot* nel tessile è, allo stato attuale, prevista per mansioni di raccordo tra segmenti successivi di lavorazione: carico-scarico e traslazione di materiali. Esistono carrelli di stesura dotati di microprocessori per il carico delle pezze e lo svolgimento automatico delle operazioni. È inoltre possibile inserire la marcatura contemporanea dei singoli particolari che verranno tagliati. Il più grosso vincolo all'introduzione di sistemi di questo tipo è costituito dalle caratteristiche dei prodotti da movimentare: la flessibilità dei tessuti causa infatti notevoli difficoltà.

Un'ultima serie di innovazioni riguarda la gestione automatica dei *magazzini* per le materie prime e per i prodotti finiti. Il caso più noto è quello del magazzino robotizzato ideato dalla COMAU per Benetton [9]. L'obiettivo è quello di consentire la gestione ordinaria e senza errori di un flusso crescente di merci da una rete produttiva ad una rete commerciale, entrambe in crescita sia in termini di dimensioni che di complessità. Il magazzino accoglie merci in uscita dagli stabilimenti Benetton e immagazzina i singoli colli, di cui è in ogni momento possibile conoscere l'ubicazione ed effettuarne il prelievo automatizzato. Il sistema gestionale infine elabora il piano di spedizione sulla base delle richieste di consegna merci da parte dei clienti e dà al magazzino l'ordine di uscita delle merci.

Numerosi vantaggi derivano dalla consistente introduzione di innovazioni informatiche - l'incremento della flessibilità produttiva, la possibilità di passare rapidamente da un prodotto all'altro e di trattare lotti di produzione molto ridotti, la diminuzione della struttura dei costi e la più rapida e completa conoscenza dell'andamento del ciclo produttivo.

È inoltre possibile identificare gli ostacoli principali incontrati dalle imprese intervistate nell'ambito delle loro strategie di rinnovamento tecnologico:

- la *scarsa standardizzazione dei prodotti* che rende necessario creare soluzioni diverse e estremamente flessibili per ogni problema di automazione e la difficoltà dei prodotti da trattare soprattutto per quanto riguarda la manipolazione e la stesura;
- la *frammentarietà del mercato* dei macchinari tessili: l'estrema varietà delle lavorazioni e l'esistenza di un universo di imprese molto diverse tra loro impedisce la formazione di un'offerta di soluzioni standard;
- la necessità di ricorrere molto spesso a soluzioni ad hoc comporta *prezzi molto elevati,* fuori dalla portata della maggior parte delle imprese;
- il fatto che non esiste una diffusa *cultura informatica* nel settore, che impedisce un costruttivo dialogo con i fornitori di tecnologie per una migliore soddisfazione delle esigenze specifiche;
- la carenza infine della *qualificazione professionale* degli addetti.

2.2 La gestione dell'informazione

Le nuove tecnologie dell'informazione assumono un ruolo centrale nell'attività di gestione delle imprese e di integrazione di queste a monte, con i propri fornitori, e a valle, con il mercato. In un settore quale il tessile/abbigliamento, caratterizzato da una struttura verticale fortemente disintegrata ed estremamente sensibile ai mutamenti di mercato, la gestione dell'informazione interna ed esterna, per quanto riguarda i legami con fornitori e clienti, rappresenta un nodo centrale del sistema aziendale ed è ritenuta l'area in cui maggiori devono assere gli sforzi di innovazione.

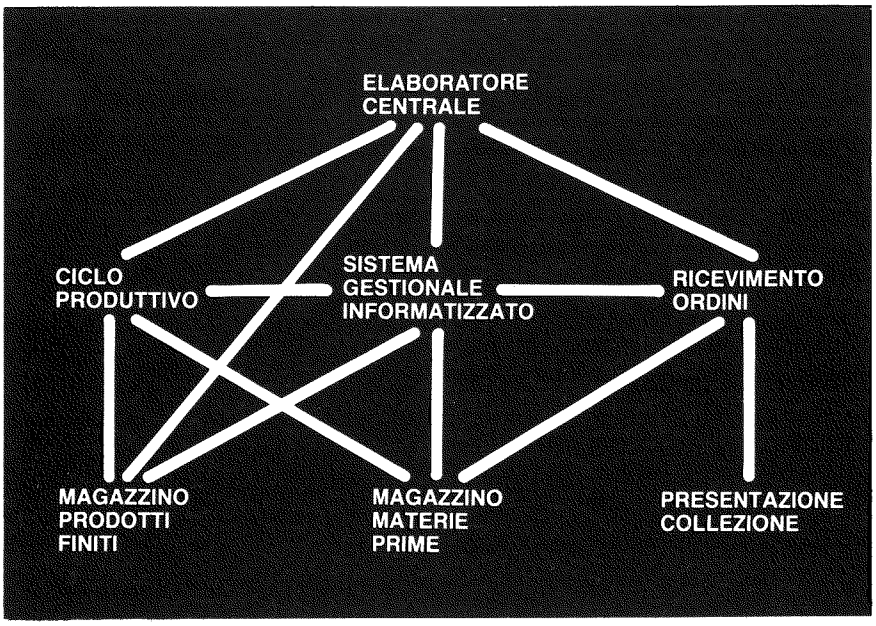
Il sistema logistico adottato dalle imprese più avanzate del settore (6 imprese del campione) integra le varie operazioni all'interno delle singole funzioni e le diverse funzioni tra loro, collegandole con il sistema informativo. L'elevato grado di integrazione e coordinamento è supportato inoltre da un sistema gestionale informatizzato che effettua piani, verifiche, proposte, ordina la successione dei lavori e coordina gli spostamenti di merci e informazioni tra i settori aziendali.

Un sistema logistico come quello descritto in figura 2 permette l'integrazione dei dati che provengono dal ciclo produttivo con quelli provenienti dal mercato, attraverso i piani di lavoro elaborati dal sistema gestionale.

La rapidità e la elasticità del ciclo logistico consentono in taluni casi l'inizio della produzione quando non si è ancora conclusa la fase di vendita (collezioni di riempimento e di riassorbimento della Benetton). La società veneta è nota anche per un'altra innovazione legata alle caratteristiche del suo ciclo logistico: la colorazione a capo finito e la conseguente spedizione dei capi in base alle richieste del mercato.

Si tratta in pratica di strategie che negli U.S.A. sono definite "Quick Response" e si ritengono fondamentali nella battaglia commerciale con i paesi emergenti. Per "Quick Response" si intende qualcosa di molto simile alla produzione "just-in-time" adottata in altri settori in-

Fig. 2 - Un esempio di sistema logistico.



dustriali, che dovrebbe condurre alla compressione dei tempi del ciclo produttivo ed all'offerta di un prodotto maggiormente rispondente alle esigenze della domanda [23].

L'obiettivo è quello di un sistema logistico (scorte, risorse, prodotti finiti) in grado di "girare", comandato dalla rete informatica, esattamente come il mercato.

2.3 Le reti informative

Uno dei casi più noti di adozione di rete telematica nel tessile/abbigliamento è quello della Benetton che nel 1978 ha istituito una rete privata di collegamento tra sede centrale e sistema di vendita [20].

Nel 1985 tuttavia si è deciso di abbandonare la soluzione adottata poichè troppo costosa e non corrispondente alla generale filosofia di flessibilità del gruppo. La nuova rete ideata si basa su criteri diversi dalla precedente: utilizzo dei servizi offerti dalle reti "multi-client" esistenti e approccio cooperativo nella definizione dei flussi di informazioni tra "headquarter" e periferia. I vantaggi derivanti dalla ristrutturazione della rete sono: in primo luogo, una maggiore flessibilità nell'espansione dei servizi; in secondo luogo, un costo proporzionato al reale utilizzo, caratterizzato da punte stagionali; infine un minore controllo da parte del centro sull'attività degli agenti.

L'introduzione della prima rete aveva infatti causato la nascita di notevoli problemi con i rivenditori, abituati a considerarsi come dei piccoli imprenditori indipendenti, e quindi poco propensi a fornire informazioni sull'andamento della propria attività. Un mutamento della filosofia aziendale verso una gestione meno esclusiva dell'informazione era perciò una necessità improrogabile di fronte ad un crescente rifiuto a inserire dati nella rete da parte dei venditori.

Diversi sono gli obiettivi che il Gruppo Tessile Miroglio si propone di raggiungere per mezzo della propria rete telematica: in primo luogo il coordinamento dei processi di produzione nei diversi impianti localizzati anche all'estero [(2)], in secondo luogo la comunicazione con il mercato finale ed infine la pianificazione dell'attività di investimento [25]. L'informazione assume perciò un ruolo cruciale per la vita dell'impresa e rappresenta un supporto essenziale per ogni decisione strategica. È importante sottolineare che il know-how necessario per trattare i problemi informativi è stato sviluppato internamente e la maggior parte delle soluzioni tecnologiche adottate sono state ideate all'interno dell'impresa. La risorsa informazione è quindi gestita a livello centrale, basandosi su un sistema informatico accentrato composto da elaboratori centrali, terminali non intelligenti presso le filiali e terminali "mobili" presso gli agenti.

È importante osservare che nel caso GTM è stato possibile introdurre una struttura informatica completamente accentrata poichè il Gruppo dipende in misura molto meno rilevante dall'esterno, sia per quanto riguarda la commercializzazione che per quanto riguarda la produzione, rispetto al caso precedentemente analizzato della Benetton.

Accanto ad esempi di adozione di reti telematiche all'interno di grandi imprese, è interessante considerare casi di adozione come l'area di Prato, caratterizzata da un numero rilevante di imprese artigianali per la produzione della lana, di piccole dimensioni, coordinate da intermediari (impannatori) che svolgono soprattutto una funzione commerciale, di connessione tra la rete produttiva e il mercato. La complessità del sistema e la pluralità degli operatori che vi intervengono a diversi livelli della scala gerarchica, implicano una rete di informazioni particolarmente complicata, il cui controllo assume un notevole valore strategico.

Infatti la modernizzazione del siste-

(2) Il Gruppo Tessile Miroglio ha 16 impianti in Italia, di cui 6 producono confezioni per donna e i rimanenti 10 tessuti, ed inoltre 4 impianti all'estero (Tunisia, Egitto e Grecia).

ma di comunicazione favorisce l'accesso alle informazioni da parte di un pluralità di operatori. Il ruolo dell'impannatore, legato all'esclusiva disponibilità delle informazioni sui mercati, viene perciò messo in pericolo. In pratica una maggiore trasparenza del mercato sarebbe destinata a modificare profondamente gerarchie e interessi economici che sono storicamente radicati nell'area pratese e rischia quindi di incontrare ostacoli insormontabili.

2.4 L'innovazione di prodotto

La consistente introduzione di innovazioni di processo che sta avendo luogo in modo sempre più considerevole nelle imprese del tessile/abbigliamento, rappresenta anche un forte stimolo all'introduzione di innovazioni di prodotto.

Nel campo dei tessuti e delle confezioni il prodotto è infatti solo apparentemente sempre lo stesso: oltre alle infinite possibilità che derivano dall'utilizzo delle fibre chimiche, l'introduzione di innovazioni a base elettronica comporta rilevanti mutamenti in termini di qualità del prodotto e di qualità del servizio.

Per ciò che concerne la *qualità del prodotto,* attraverso l'introduzione di tecnologie informatiche si sono ottenuti notevoli progressi sia in termini di omogeneità che in termini di varietà. Una più elevata omogeneità delle caratteristiche dei prodotti è garantita dai sofisticati sistemi per il controllo di qualità e dall'automazione di fasi di produzione, quali la tintura e la stampa, il taglio ed il cucito. La maggiore varietà dei disegni e delle composizioni cromatiche dei tessuti, nonchè dei modelli dei capi confezionati, è invece la conseguenza dell'introduzione di sistemi come il CAD e il CAD/CAM che consentono una maggiore creatività e permettono di produrre un campionario più ampio, a costi contenuti.

Infine, notevoli miglioramenti intervengono in termini di *qualità del servizio*, ritenuta da molte imprese un fattore strategico molto importante sul quale puntare nel gioco competitivo. Miglioramenti della qualità del servizio possono riguardare l'offerta di capi maggiormente rispondenti ai gusti del consumatore, resa possibile dalla più ampia disponibilità di informazioni sul mercato, dall'esistenza di canali di interazione con la domanda particolarmente sviluppati e inoltre dalla possibilità di trasferire rapidamente i segnali che provengono dall'esterno ai sistemi di produzione.

Per quanto riguarda la commercializzazione, una distribuzione più capillare ed un servizio uniforme sono garantiti da nuove forme di accordo commerciale adottate soprattutto dalle imprese di confezioni, quali per esempio il franchising. La formula del franchising, senz'altro una delle strategie vincenti della Benetton, permette infatti alle imprese la costruzione di una rete commerciale molto estesa, omogenea in termini di immagine e finanziariamente poco costose.
Infine un tipo di servizio molto particolare viene offerto dal Gruppo Finanziario Tessile: si tratta della linea "vestito su misura". In un numero limitato di negozi il GFT offre infatti ai suoi clienti un servizio di sartoria su misura in 48 ore. La rapidità di trasferimento delle informazioni e l'automazione dei sistemi di produzione hanno in questo caso un ruolo cruciale, poichè solo una perfetta coordinazione delle diverse fasi del sistema rendono possibile all'impresa offrire un servizio così particolare, in tempi tanto brevi.

3. L'innovazione macro-organizzativa

Secondo Williamson [27, 28, 29] le attività economiche si svolgono all'interno di organizzazioni gerarchiche o sul mercato, a seconda del peso relativo dei costi di organizzazione e di quelli di transazione. Le transazioni possono differire fra loro in termini di specificità dell'oggetto scambiato, di incertezza e di frequenza. Il mercato è la forma più idonea in situazioni di certezza, non specificità dell'oggetto scambiato e occasionalità delle transazioni, mentre risulta più efficace una organizzazione gerarchica quando vi è incertezza, quando il prodotto non è standardizzato e quanto più le transazioni sono ricorrenti [22].

Le tecnologie dell'informazione possono influenzare la scelta tra mercato e impresa integrata in seguito all'impatto che esse hanno sui costi di transazione e di organizzazione. In particolare, le nuove tecnologie rendono meno definiti i confini tra mercato e impresa, generando una ampia gamma di situazioni intermedie *non market - non hierarchy,* quali relazioni contrattuali, joint ventures, franchising, etc. [3].

Una immagine sintetica della struttura macrorganizzativa del settore tessile/abbigliamento, è fornita dallo schema proposto nella figura 3. Nella cornice superiore della tabella è indicato uno spettro continuo di forme organizzative che va dalla completa integrazione verticale ("make") al mercato ("buy"), passando attraverso numerose forme intermedie di accordo ("make together"); sulle righe invece sono indicate le funzioni che compongono il ciclo produttivo tessile.

L'*integrazione verticale* di tutte le fasi del ciclo, come avviene tradizionalmente nei grandi gruppi, sembrerebbe sempre meno diffusa poichè impone notevoli rigidità dal punto di vista produttivo. Tuttavia, vi è motivo di ritenere che alcuni aspetti delle tecnologie informatiche spingano verso una maggiore integrazione verticale. Alcune funzioni svolte precedentemente in modo separato infatti risultano, in seguito all'adozione di nuovi sistemi tecnologici, tra loro necessariamente integrate. Per esempio, l'adozione di sistemi CAD/CAM implica l'integrazione della funzione di progettazione con quella di produzione, di tintoria o di sviluppo delle taglie e "ingegnerizzazione" dell'idea creativa. Un'ulteriore spinta all'integrazione deriva dall'adozione di sistemi informativi che connettono la funzione produttiva con quella gestionale e quella commerciale, per cui le indicazioni di produzione vengono ad essere strettamente legate alla pianificazione strategica e alle politiche di marketing.

Anche per quanto riguarda le soluzioni tecnologiche adottate dalle imprese del settore, le realizzazioni "in house" sono molto frequenti sia

in termini di completo sviluppo interno della tecnologia, sia di adeguamento interno di soluzioni standard, sia infine di collaborazione con consulenti esterni. La necessità di sviluppare internamente i sistemi tecnologici adottati deriva prevalentemente dalle difficoltà di reperire sul mercato prodotti standard, soprattutto per le fasi del ciclo più specializzate e meno diffuse.

Vi è infine un'ultima motivazione della spinta verso l'integrazione verticale: l'aumentata disponibilità di informazioni in tempo reale sull'andamento del ciclo produttivo. Le nuove tecnologie informatiche consentono infatti un maggiore controllo sulle diverse fasi che compongono il ciclo.

Spostandoci verso destra lungo la cornice superiore del nostro schema troviamo gli *equity agreements,* forme di accordo che prevedono una partecipazione azionaria tra i partner. Si possono definire "equity agreements" le *joint ventures* che talvolta sorgono tra imprese di confezioni e grandi creatori di moda. I servizi creativi forniti dagli stilisti, oltre che la possibilità di usare un nome conosciuto per i prodotti, si associano così alle capacità produttive e organizzative e alla disponibilità di una rete commerciale delle imprese confezioniste per dar vita a nuove società.

La creazione di società avviene anche per la produzione di tecnologie di produzione da vendere all'esterno, spesso ai paesi emergenti, realizzate sfruttando il know-how sviluppatosi per risolvere autonomamente i problemi interni.

Gli accordi che non prevedono partecipazione azionaria da parte dei partner sono definiti *non equity agreements.*

Nel nostro settore abbastanza diffusi sono i casi di *partecipazione finanziaria* del partner dominante *per l'acquisto della tecnologia.*

Rientrano nella categoria dei "non equity agreements" anche gli accordi che prevedono forme di *controllo* dell'impresa dominante sui partner. I più diffusi sono sicuramente gli *accordi di subfornitura* che consistono nel far realizzare all'esterno alcune fasi di trasformazione del prodotto che poi è assemblato e terminato all'interno.

Esistono diverse forme di contratto: talvolta sono fornite solamente le specifiche di produzione, talaltra vengono dati i modelli da realizzare e in alcuni casi infine sono fornite anche le materie prime.

In campo commerciale, il *franchising* può rientrare nelle forme di controllo in quanto lega i proprietari dei negozi che appartengono alla rete commerciale con accordi contrattuali che impongono l'esclusività, fissando i prezzi e indicando gli standard dei servizi da garantire ai clienti.

Altre forme di "non equity agreements" sono gli accordi formali di *collaborazione tra imprese e i consorzi* per lo svolgimento di alcune attività in comune (Citer, ICE, Enea).

Il passo successivo verso il mercato sul nostro schema conduce alle *aree sistema,* caratterizzate da accordi informali e da bassi costi di transazione legati alla frequenza dei contatti, alla disponibilità di informazioni e alla consuetudine dei rapporti. In realtà caratterizzate da un sistema complessivo di aziende piccole, medie e artigianali, produttive e di servizi, che lavorano tutte nel settore tessile con infinite interrelazioni fra loro, il fattore di produzione *comunicazione-informazione* ha un ruolo importantissimo; di conseguenza interventi di tipo telematico sono attualmente in corso a Prato e Como.

I progetti di sviluppo tecnologico a livello di aree sistema spesso riguardano anche le innovazioni di processo e, oltre a iniziative a carattere pubblico o semi-pubblico, l'area sistema è spesso sede di fornitori di tecnologia. La presenza delle imprese clienti e la possibilità di interagire maggiormente con esse, comporta frequentemente la nascita di produttori di tecnologie molto specializzate che offrono prodotti fortemente "custom-oriented". Il distretto diviene quindi "distretto integrato", caratterizzato da una attività leader che a sua volta induce lo sviluppo di attività complementari specializzate di produzione e/o servizio.

Da un punto di vista commerciale, infine, il distretto svolge una funzione di "immagine" molto importante per le imprese che vi appartengono. La concentrazione territoriale e l'identificazione del prodotto con specifiche caratteristiche di qualità garantite dall'area attraggono i clienti e rappresentano uno strumento commerciale estremamente efficace.

Passando in ultimo a considerare il *mercato,* vediamo come tutte le funzioni delle imprese possono essere realizzate semplicemente acquistando all'esterno servizi o prodotti. Sul mercato è possibile infatti acquistare servizi fieristici per la presentazione delle collezioni, come pure è possibile vendere i propri prodotti per mezzo di rappresentanti a grossisti e dettaglianti. Anche le innovazioni tecnologiche possono essere acquistate sul mercato presso produttori, sia in quanto prodotti standard che in quanto soluzioni ad hoc. Infine, sul mercato naturalmente è possibile acquistare servizi offerti dagli stilisti, quindi modelli e idee creative nonchè prodotti finiti. Talvolta infatti le imprese si limitano a commercializzare con il proprio marchio modelli ideati e confezionati interamente all'esterno, presso imprese che svolgono tutte le fasi del ciclo produttivo o a loro volta si avvalgono di terzisti esterni.

Dall'analisi svolta sembra possibile sostenere l'esistenza di una tendenza attuale verso la concentrazione delle funzioni strategiche. Considerando lo schema sintetico da noi proposto (fig. 3), si può ipotizzare uno spostamento verso sinistra, quindi verso maggiori gradi di controllo del ciclo se non verso l'*integrazione verticale,* almeno nel prossimo futuro. Appare evidente infatti la centralizzazione delle decisioni riguardo gli aspetti commerciali, la pianificazione strategica, la progettazione dei prodotti, l'adozione di tecnologie informatiche e non.

Lo spostamento verso l'alto del baricentro strategico è dovuto alla quantità e qualità di informazioni su tutti gli aspetti del sistema economico dell'impresa, accessibili alle funzioni gerarchiche più elevate.

Le tecnologie informatiche sembrano quindi influire in misura maggiore sulla riduzione dei costi di orga-

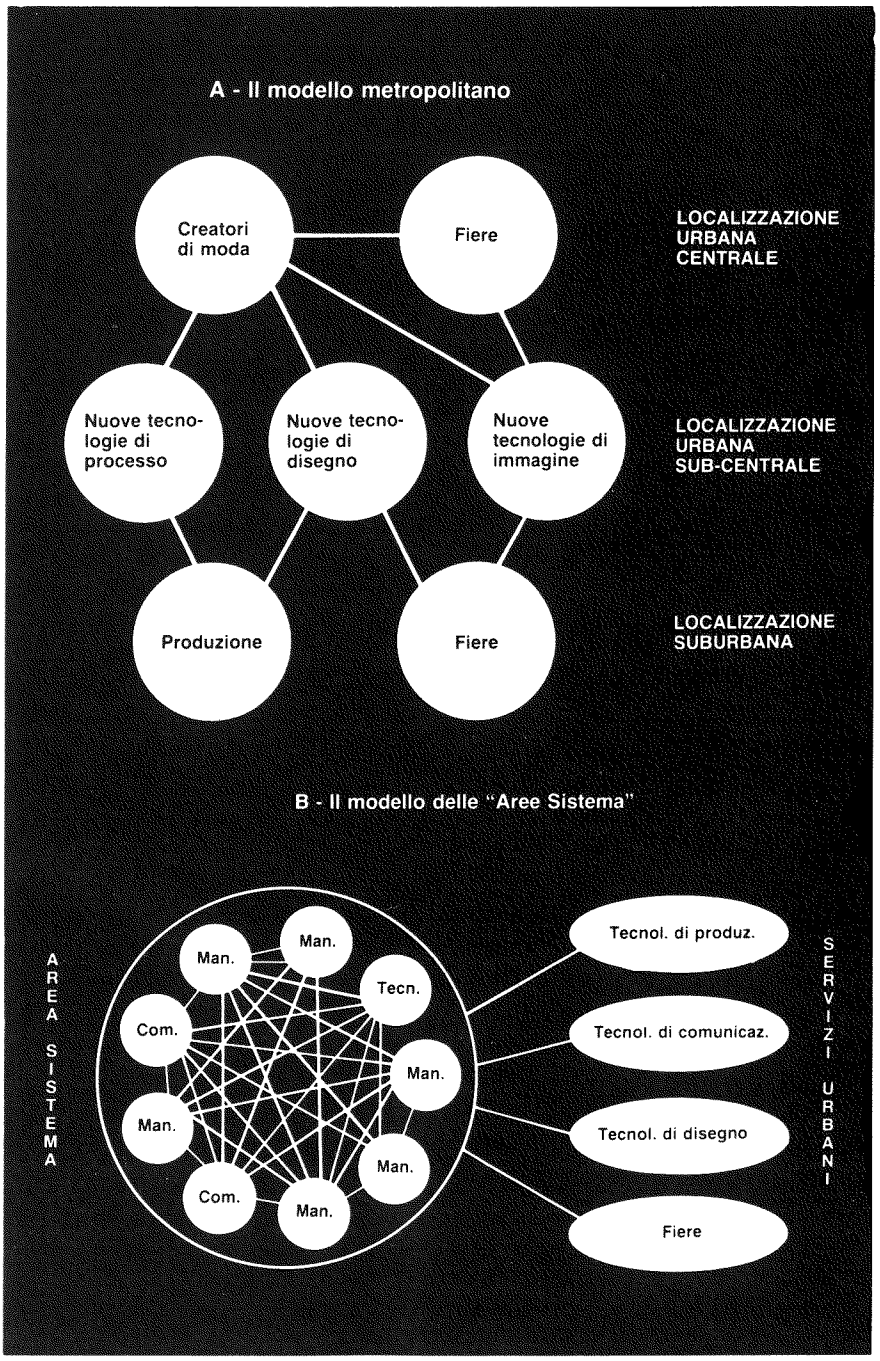
Fig. 3 - La struttura macrorganizzativa del settore tessile-abbigliamento.

Forma organizzativa / Funzione	MAKE	MAKE TOGETHER					BUY
	Integrazione verticale	equity agreement	Non equity agreement			Aree sistema	Mercato
			Partec. finanz. tecnologia	Controllo	Accordi, collab., consorzi		
FIERA					rapporti tra fiere e aree specializz. (MI, CO, FI).		acquisto servizi fieristici
MODA	stilisti interni	joint-venture con stilisti			banche dati disegni e immagini (CITER)		acquisto servizi stilisti
TELEMATICA	realizzazioni in-house			reti private di fornitori o punti di vendita	servizi telematici offerti da consorzi es.: ICE, CITER	aree cablate es.: Prato, Como	acquisto servizi telematici
COMMERCIALIZZAZIONE	commercializ. interna		franchising: es.: Benetton	franchising		distretto come "immagine"	commercializ. esterna
TECNOLOGIE DI PROCESSO	realizzazioni in-house	creazione società per produrre tecnologia, es.: GFT			collab. sviluppo progetti tecnol.; es.: ENEA/Como...	produttori di tecnol. nell'area: distretti integr.	acq. serv. tecn.: magaz. Benetton, GFT, Loropiana
PRODUZIONE	produzione interna		fornitura tecnologie ai terzisti	controllo subfornitori		terzisti special. nelle diverse fasi del ciclo	acquisto prodotti da comm. sul mercato
FORMAZIONE	formazione interna				formazione gestita da consorzi	formazione specialistica aree sistema	corsi offerti dal mercato (es.: produt. di tecn.)

nizzazione piuttosto che su quelli di transazione, in quanto, almeno in un primo periodo, sono di più facile introduzione in realtà caratterizzate da gerarchie di potere consolidate, riconoscibili e accettate. Inoltre la difficoltà e complessità di adozione nonchè il costo spesso molto elevato delle nuove tecnologie dell'informazione rappresentano un ostacolo alla loro diffusione nelle imprese di piccola dimensione e confermano la natura "conservatrice" del modello di diffusione delle innovazioni.

Tuttavia, sebbene a livelli gerarchici elevati non vi sia alcuna tendenza al decentramento decisionale, esistono varie *forme intermedie di non-organizzazione* (o di "quasi-organizzazione") dell'industria che vanno assumendo un ruolo sempre più importante come alternativa al mercato e alla completa integrazione verticale, in particolari fasi del ciclo.
La necessità di rivolgersi all'esterno, la ricerca di sinergie e di complementarietà tra i partner sono in parte legate all'incremento di complessità e di specializzazione dei prodotti e dei mercati. In primo luogo, la tecnologia diviene sempre più complessa e sofisticata e la sua utilizzazione richiede consulenze di esperti; in secondo luogo, la ricerca della varietà e dell'originalità dei modelli comporta spesso il ricorso a servizi creativi esterni; in terzo luogo, anche la commercializzazione dei prodotti talvolta si avvale di accordi con l'esterno che permettono l'espansione della rete con investimenti limitati; infine alcune fasi del ciclo

Fig. 4 - Struttura e geografia del settore tessile-abbigliamento.



produttivo sono spesso decentrate presso unità produttive esterne per motivi di costo e flessibilità.

L'esistenza di una "filiera" interamente sviluppata dalla produzione alla commercializzazione e alla "fashion creation", la possibilità quindi di accedere a servizi specializzati, produzioni complementari e ad una rete di subfornitori presso cui decentrare alcune fasi del ciclo in un rapporto di cooperazione stretta, assume, secondo noi, una importanza fondamentale tra le motivazioni del successo del tessile/abbigliamento italiano.

Nello scenario futuro che noi riteniamo possibile è identificabile una tendenza verso l'integrazione verticale per quanto riguarda le funzioni strategiche e di controllo e quindi verso un rafforzamento della grande impresa, nonchè verso una sempre maggiore diffusione di forme intermedie di non-organizzazione per quanto riguarda il ricorso a servizi e produzioni esterne. Il ruolo assunto dall'ambiente esterno risulta essere quindi molto importante, anche se al distretto industriale nel senso tradizionale del termine si sostituisce il *distretto industriale allargato,* inteso dal punto di vista territoriale in modo più ampio. In altre parole, noi crediamo che ai modelli tradizionali della grande impresa integrata, quasi "automatica", e della piccola e media impresa specializzata nella singola fase del ciclo, si vada sostituendo, da una parte, un nuovo tipo di grande impresa "tentacolare", fortemente accentrata nelle sue funzioni strategiche e con numerose diramazioni, e un nuovo tipo di piccola impresa, integrata in una rete locale con altre piccole imprese, all'interno di un sistema di rapporti di forza in continua evoluzione che regolano sia la dinamica innovativa che la "appropriazione" di risultati economici fra partner (figura 4).

Bibliografia

[1] ANTONELLI C., COZZI G., FORNENGO G., GAMBARDELLA A., RULLANI E., ZANFEI A., 1986, *Structural impacts of telematics on automobile and textile & clothing industries,* Fast Final Report, luglio, Milano.

[2] ARCANGELI F., CAMAGNI R., 1987, "Sistemi tecnologici e diffusione territoriale", relazione presentata al Convegno su *Mutazioni tecnologiche e condizionamenti internazionali: reazioni del sistema delle piccole e medie imprese italiane,* Padova, 28-29 maggio.

[3] BALCET G., VIESTI G., 1986, "Fra mercato e gerarchia: alcune riflessioni sugli accordi tra imprese", *Economia e politica industriale,* n. 49, marzo, pp. 39-64.

[4] BECATTINI G., 1979, "Dal settore industriale al distretto industriale. Alcune considerazioni sull'unità di indagine dell'econmomia industriale", *Rivista di Economia e Politica Industriale,* n. 1, pp. 7.21.

[5] CAMAGNI R., 1985, "Spatial diffusion of pervasive process innovation", *Papers and Proceedings of the Regional Science Association,* n. 54.

[6] CAMAGNI R., 1986, "New technologies as a response to the crisis in the Italian North-West", in Federwisch J., Zoller H.G., eds., 1986, *Téchnologie nouvelle et rupturers régionales,* Economica, Paris.

[7] CAMAGNI R.,1987, "The spatial implications of technological diffusion and economic restructuring in Europe with special reference to the Italian case", paper presentato all'*Expert Meeting* su "Technological Developments and Urban Change", OCDE, Parigi, 29-30 giugno.

[8] COASE R.H., 1937, "The nature of the firm", *Economica,* november, pp. 386-403.

[9] COPPOLA B., CROCE B., 1985, "Benetton, modello veneto e tecnologie dell'informazione: il magazzino automatizzato", *Lito-Newsletter,* n. 6.

[10] DEI OTTATI G., 1986, "Distretto industriale, problemi delle transazioni e mercato comunitario: prime considerazioni", *Economia e politica industriale,* n. 51, pp. 93-121.

[11] DUGNANI G., 1987, "CAD tessile", *Tinctoria,* n. 3, marzo.

[12] HOFFMAN K., 1985, "Clothing, chips and competitive advantage: the impact of microelectronics on trade and production in the garment industry", *World Development,* vol. 13, n. 3, pp. 371-392.

[13] MARIOTTI S., 1982, *Efficienza e struttura economica: il caso tessile/abbigliamento,* Franco Angeli, Milano.

[14] MARIOTTI S., CAINARCA G.C., 1986, "The evolution of transaction governance in the textile-clothing industry", *Journal of Economic Behaviour and Organization.*

[15] MAZZONIS D., 1985, "A project for innovation in Prato", paper presentato a *International Workshop* su "Integration of new technologies in traditional sectors" San Miniato, 27-30 novembre.

[16] MARITI P., SMILEY R.H., 1982, "Accordi di collaborazione tra imprese in una economia industriale moderna", *Economia e politica industriale,* n. 33.

[17] MONACO T., 1987, "Le modalità microeconomiche del controllo sulle risorse", in Camagni R., Predetti A. *La Struttura Economica - Progetto Milano,* Franco Angeli Editore, Milano, in via di pubblicazione.

[18] RESEAU, 1985, *Scenario telematico e territorio. Lettura di una esperienza in corso,* Rapporto di ricerca, ottobre, Milano.

[19] RESEAU, 1986, *Progetto telematica per la provincia di Como,* Rapporto di ricerca, settembre, Milano.

[20] RULLANI E., ZANFEI A., 1984, "Benetton: invenzione e consolidamento di un sistema internazionale", *Bollettino Ospri,* n. 1.

[21] SANTANGELO A., 1986, "CAD/CAM nell'abbigliamento", *Lito Newsletter,* n. 1, pp. 5-8.

[22] SILVA F., 1985, "Qualcosa di nuovo nella teoria dell'impresa?", *Economia e politica industriale,* n. 1, aprile.

[23] Textile World, 1987, *Special Report: Quick Response,* february.

[24] U.S. Congress, 1987, *"The U.S. Textile and Apparel Industry,* Special Report, april, Washington.

[25] ZANFEI A., 1984, "Gruppo Tessile Miroglio: internazionalizzazione e ridefinizione del livello tecnologico e qualitativo", *Bollettino Ospri,* n. 2, dicembre.

[26] ZANFEI A., 1985, "Cambiamento tecnologico e strategie di internazionalizzazione delle imprese italiane", *Economia e politica industriale,* n. 46, pp. 93-132 e n. 47, pp. 97-136.

[27] WILLIAMSON O. E., 1967, "Hierarchical control and optimum firm size", *The Journal of Political Economy,* april, n. 2, pp. 123-138.

[28] WILLIAMSON O. E., 1975, *Markets and Hierarchies: Analysis and Antitrust Implications,* The Free Press, New York.

[29] WILLIAMSON O. E., 1983, "Vertical integration and related variations on a transaction-cost economics theme", in Stiglitz J.E., Mathewson G.F., eds., 1983, *New developments in the analysis of market structure,* London, Macmillan.

Le Database Machine

LEONARDO FELICIAN
ANDREA PAIUSCO

Dipartimento di Matematica e Informatica
Università di Udine

1. Introduzione

La necessità di attribuire una definizione chiara ed esaustiva al termine *database machine* si scontra con la varietà di soluzioni ed approcci architetturali proposti in questo particolare settore.

In prima istanza si può dire che una *database machine* non è altro che un'interfaccia tra l'unità centrale e le memorie di massa (*backend*), che si fa carico delle funzioni di gestione della base di dati e delle routine di I/O normalmente proprie del computer principale, detto solitamente ospite (*host*), per mezzo di hardware e software specializzato, Da un punto di vista qualitativo, questa definizione è abbastanza soddisfacente, poichè lascia intravedere i possibili vantaggi offerti in termini di prestazioni dall'introduzione di una *database machine;* dal punto di vista tecnico tuttavia, essa appare alquanto insufficiente. Si impone quindi l'introduzione di una classificazione più precisa delle *database machine* [8].

L'accesso convenzionale ai dati generalmente in uso nei sistemi per la gestione di basi di dati è rappresentato in fig. 1.

Una prima proposta che si differenzia da questo schema consiste nell'inserimento come *backend* tra il computer ospite e il controllore dei dischi di un altro computer, solitamente un mini, a cui demandare tanto la gestione della base di dati quanto le attività di I/O. Lo schema risultante è rappresentato in fig. 2.

Questo schema non comporta praticamente nessun intervento di tipo hardware: al contrario, l'innovazione rispetto all'approccio tradizionale è esclusivamente a livello della realizzazione software del sistema per la gestione di basi di dati. Sulla scorta di questa osservazione non stupisce il nome di *single-backend software* attribuito ad architetture di questo tipo. Una simile *database machine* rappresenta, per così dire, una soluzione statica, poichè l'eventuale crescita della base di dati ne penalizza fortemente le prestazioni.

L'idea di replicare sia i singoli *backend* che il software su di essi residente ha condotto ad un approccio definito *multiple-backend software,* illustrato in fig. 3. Anche in questo caso, l'aspetto hardware riveste un'importanza secondaria.

Al contrario, l'utilizzo di un *backend* speciale (usualmente chiamato *intelligent-controller*) rientra in un ambito architetturale di tipo marcatamente hardware. In questo caso, infatti, il *controller* dei dischi viene realizzato in modo opportuno per mezzo di microprocessori, memorie e linee di comunicazione, con l'evidente scopo di migliorare l'efficienza delle operazioni di accesso alla base di dati, tramite indirizzamenti per valore. Tuttavia, il linguaggio di manipolazione dei dati (*Data Manipulation Language: DML*) è normalmente supportato dal computer ospite che si fa inoltre carico della sicurezza e dell'integrità della base di dati. Questi principi architetturali sono schematizzati in fig. 4. Esempi di *database machine* di questo tipo sono, in particolare, i prodotti commerciali IDM 500 della Britton-Lee e i DBP della Intel.

Il naturale sviluppo di una simile architettura consiste nel realizzare un vero e proprio computer che permetta, oltre ad un appropriato controllo delle unità a disco, una gestione ottimale e completa della base di dati; in particolare questo approccio, denominato *hardware-backend,* si differenzia da tutti i precedenti poichè generalmente comporta modifiche e innovazioni all'attuale tecnologia hardware, in particolare per la progettazione di processori che eseguano operazioni proprie dell'algebra relazionale. Lo schema di fig. 5, relativo a quest'ultimo caso, è una semplice evoluzione del precedente.

Questa prima classificazione delle *database machine,* seppure semplice e lineare nel suo sviluppo, non deve far pensare che categorizzazioni così precise siano sempre possibili nei casi concreti: in realtà la distinzione fra *backend software* o *hardware* può essere talvolta artificiosa.

2. Le memorie associative

Il concetto fondamentale su cui si basa qualsiasi *database machine* è quello di accedere ai dati per valore, piuttosto che per indirizzo; questa proprietà sottintende l'utilizzo di memorie associative. Per non generare confusione è necessario chiarire

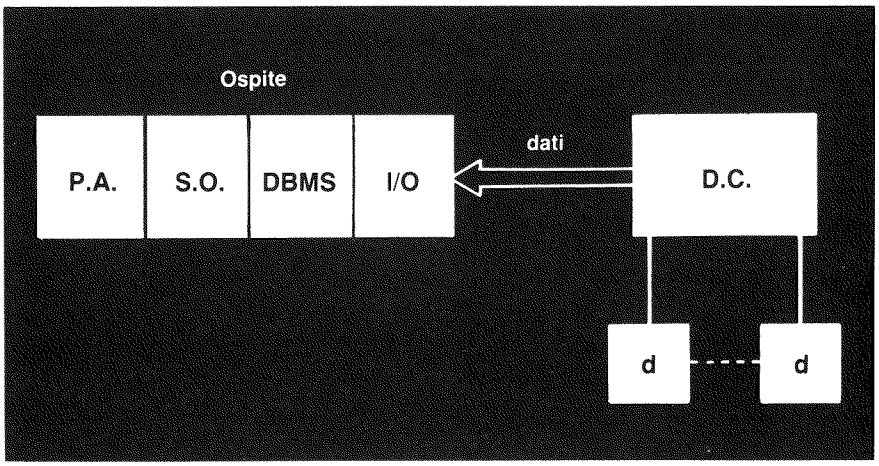


Fig. 1 - Schema di accesso ai dati in un sistema per la gestione di basi di dati.
Legenda:
P.A.: programmi applicativi
S.O.: sistema operativo
DBMS.: data base management system
D.C.: disk controller
d.: disco

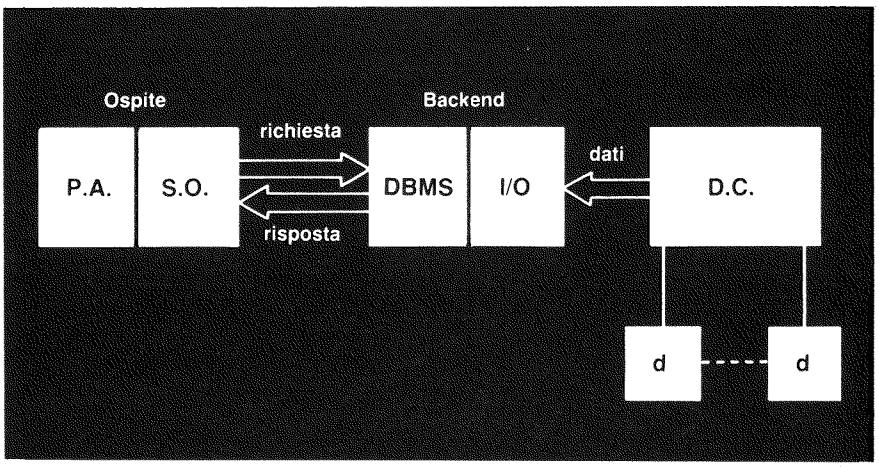


Fig. 2 - Inserimento di un computer *backend* dedicato alla gestione dei dati.

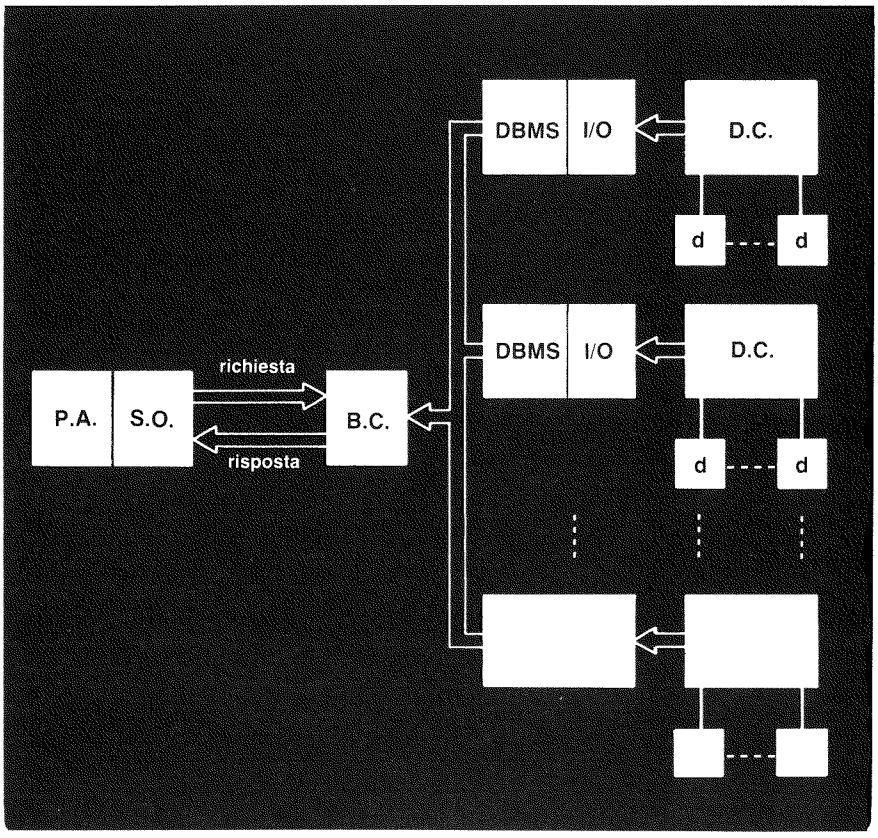


Fig. 3 - Architettura con *multiple-backend.* (B.C. = *backend controller).*

Fig. 4 - Architettura con *intelligent-controller.* (I.C. = *intelligent controller).*

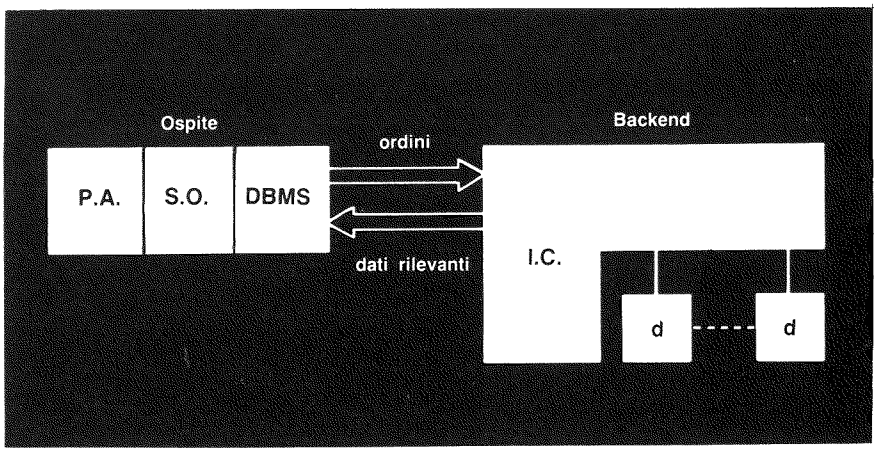
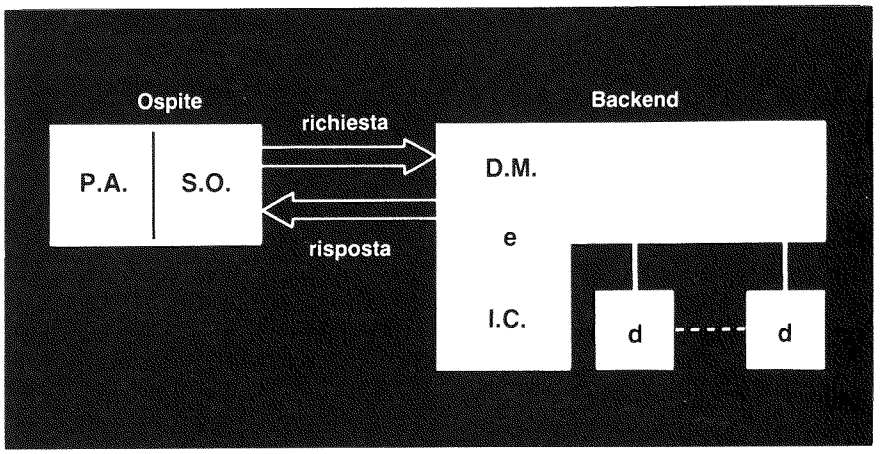


Fig. 5 - Architettura *hardware-backend.* (D.M. = *database management).*



che, propriamente, con il termine memoria associativa (AM) viene indicata un'unità di memoria di poche migliaia di byte nella quale è possibile eseguire una ricerca parallela. In pratica, quando un valore viene presentato su un opportuno ingresso di una AM, questo viene confrontato simultaneamente con il contenuto, o con una sua parte, di tutte le celle; il contenuto della cella il cui confronto ha dato esito positivo, viene letto o modificato con il valore di input, in un tempo di accesso che è tipicamente di 50-100 nsec [11]. In realtà,l'utilizzo di memorie di questo tipo è gravemente penalizzato dal loro costo; pur potendo ipotizzare una futura proliferazione di *database machine* implementate mediante hardware associativo, architetture di questo tipo sono, a tutt'oggi, estremamente rare (si veda STARAN nel prossimo paragrafo).

Il concetto di memoria associativa è stato tuttavia esteso, per indicare una generica memoria su cui venga eseguita una ricerca seriale di dati che corrispondano a determinati criteri. È questo tipo di dispositivo che viene in pratica utilizzato nella maggior parte delle *database machine.*

3. Una soluzione di tipo software

In questo paragrafo si vuole accennare a come la progettazione di una *database machine* possa essere realizzata esclusivamente per mezzo di software di nuova concezione. Per far questo si farà riferimento al progetto MDBS (*Multiple-backend Data Base System*) [7] che appartiene alla categoria dei software *multiple-backend* descritta in precedenza.

È necessario innanzitutto definire il concetto fondamentale di *"cluster* di registrazioni", che trova la propria giustificazione nella struttura della *database machine* considerata. I *cluster* vengono determinati da chi crea la base di dati; egli, infatti, specifica un certo numero di descrittori, che possono essere definiti, usando una terminologia relazionale, come unioni e disgiunzioni di predicati sugli attributi (campi) dei record. Ad esempio, un descrittore potrebbe essere un'espressione del tipo seguente: (0 < stipendio < 1000000). Un *cluster* è un gruppo di registrazioni tale che ognuna di esse è derivabile dal medesimo insieme di descrittori; in questo modo la base di dati viene partizionata in classi di equivalenza, i *cluster* appunto. Se l'unico descrittore utilizzato fosse quello dell'esempio precedente, la base di dati verrebbe suddivisa in due classi: la prima costituita dalle registrazioni che possiedono un

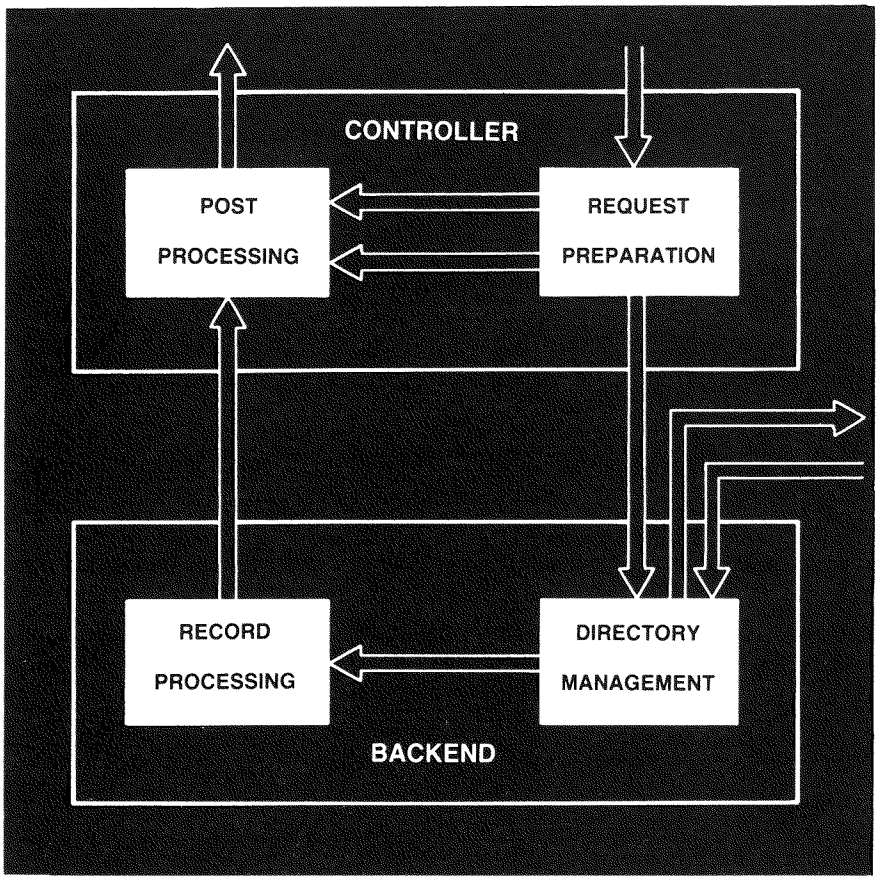


Fig. 6 - Funzioni rappresentate come moduli.

campo stipendio con valore compreso entro i limiti definiti, la seconda formata da registrazioni contenenti un campo stipendio, ma con valore non compreso, e da registrazioni prive del campo considerato.

Viene definita in questo modo una corrispondenza fra predicati, descrittori e *cluster* che consente l'organizzazione di questi dati in tabelle e, come si vedrà, l'esecuzione di elaborazioni parallele sui *backend.*

La struttura complessiva di MDBS può essere descritta mediante le funzioni proprie del *controller* e dei *backend.* Le prime sono: funzioni di preparazione delle richieste, funzioni di inserimento, funzioni di *post processing.* In particolare, le funzioni di preparazione delle richieste hanno il compito di controllare le richieste e le transazioni prima che queste vengano trasmesse ai *backend;* le funzioni di inserimento si occupano di fornire le informazioni aggiuntive richieste dai *backend* per l'esecuzione di tale operazione. Infine, le funzioni di *post processing* hanno il compito di raccogliere ed organizzare i dati forniti dai *backend* prima di trasmetterli all'ospite.

Per quel che riguarda le funzioni di ciascun *backend,* esse vengono classificate come funzioni di *directory management,* funzioni di elaborazione di registrazioni e funzioni di controllo di concorrenza.

Le prime eseguono la ricerca di descrittori e *cluster;* inoltre, esse generano gli indirizzi delle registrazioni contenute nei *cluster* considerati. Alle funzioni di elaborazione delle registrazioni sono demandati la memorizzazione, il recupero e la selezione dei record; le funzioni di controllo di concorrenza infine, eseguono le operazioni necessarie ad assicurare la consistenza dei dati in presenza di elaborazioni concorrenti.

Alcune delle funzioni considerate risultano maggiormente chiare quando si esamini in dettaglio l'esecuzione di una generica richiesta; si supporrà, per semplicità, che quest'ultima non comporti operazioni di inserimento. Dal momento che richieste diverse producono sequenze di azioni molto simili, si considererà una richiesta di *retrieve* (cfr. fig. 6).

Il modulo di *request preparation,* dopo aver ricevuto e controllato la richiesta, la invia, in ciascun *backend,* al modulo *directory management* e ne trasmette le caratteristiche al modulo *post processing;* i predicati che compongono la richiesta vengono utilizzati, suddivisi in insiemi diversi, uno per ciascun *backend,* per la ricerca dei descrittori opportuni. Una volta che il modulo *directory management* ha determinato gli identificatori dei descrittori corrispondenti all'insieme fornitogli, li trasmette agli altri *backend*; al ricevimento degli identificatori, infine, esso determina i *cluster* corrispondenti, stabilendo gli indirizzi in memoria secondaria delle registrazioni in essi contenuti. Questi dati vengono trasmessi al *record processing* che si occupa di prelevare i dati dal supporto di memorizzazione e selezionare quelli che fra loro soddisfano la richiesta. Successivamente, esso estrae i valori cercati dalle registrazioni selezionate, inviandole al modulo *post processing* nel *controller.*

Da qui esse verranno trasmesse al computer ospite.

Questa breve descrizione di un esempio di software *multiple-back-end* evidenzia come in soluzioni di questo tipo la definizione di un modello dei dati opportuno sia fondamentale; tale aspetto non è, in approcci di tipo hardware, così importante.

Tuttavia, a livello funzionale, la distinzione fra architettura hardware e software non è così marcata e ciò apparirà più chiaro in seguito. Le funzionalità individuate nei progetti sono spesso le medesime e le soluzioni utilizzate sono teoricamente intercambiabili. Di fatto, esempi di macchine "ibride" non sono infrequenti e recenti progetti mostrano una sempre maggior connessione fra problematiche HW e SW [8].

4. Alcune alternative architetturali

Si vuole a questo punto concentrare l'attenzione sugli sviluppi architetturali più significativi nel campo delle *database machine* di tipo hardware. Gli esempi considerati rappresentano un'evoluzione cominciata nella seconda metà degli anni '70. È in questo periodo, infatti, che vengono progettate nuove architetture con lo scopo di migliorare l'efficienza nell'accesso ai dati per mezzo dell'indirizzamento per valore, anzichè per posizione. Esse presentano alcune caratteristiche unitarie: la realizzazione sotto forma di *back-end*, collegato al computer ospite attraverso canali di I/O o linee di comunicazione, la comune classificazione nell'ambito degli *intelligent-controller,* l'implicita natura relazionale.

Un primo esempio di architetture di questo tipo è CAFS (*Content Addressable File Store*) [2], [3]: esso appartiene, secondo una classificazione dovuta a Bray e Feeman [2], alla categoria delle *database machine* a ricerca diretta con singolo processore; ciò significa che CAFS non utilizza alcun tipo di *buffering* e non supporta elaborazioni parallele. La fig. 7 mostra l'architettura complessiva di un sistema basato su CAFS.

La base di dati è memorizzata su una serie di dischi i cui output possono essere trasmessi contemporaneamente su una linea di tipo multiplex. CAFS provvede a ridurre il volume dei dati da passare al computer ospite, per mezzo di operazioni di tipo select e project; è chiaro come funzioni diverse da queste rimangano a carico del sistema per la gestio-

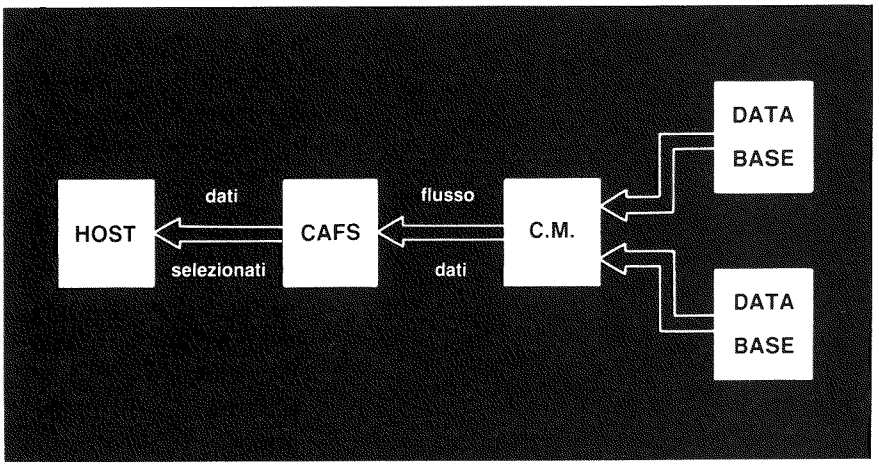


Fig. 7 - Architettura CAFS.
(C.M. = *channel multiplexer).*

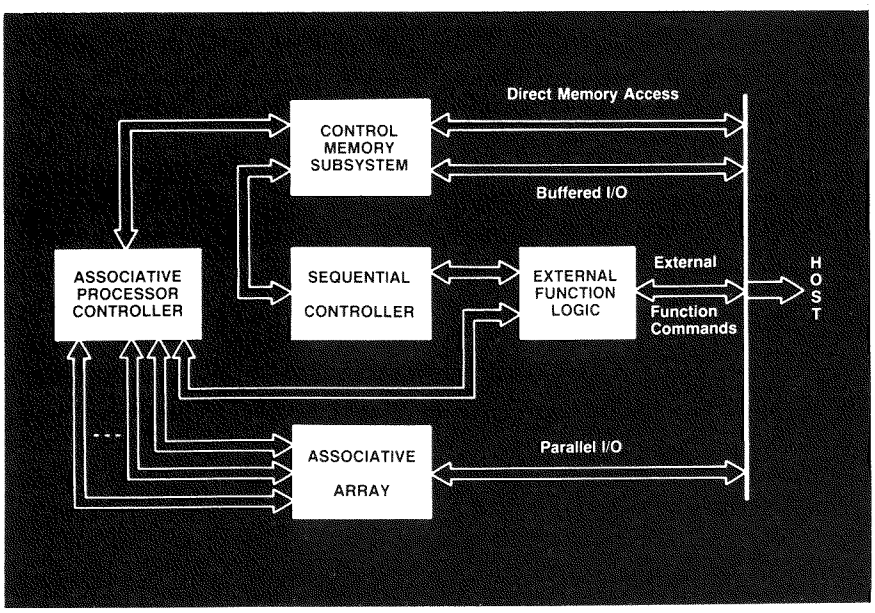


Fig. 8 - Architettura di STARAN.

ne di basi di dati residente sul computer ospite e debbano essere eseguite prima della definitiva trasmissione dei dati all'utente.

In particolare, va rilevato che CAFS costituisce esclusivamente un *retrieval system* e perciò non consente modifiche on-line dei dati.

Diversamente da CAFS, STARAN e RAP sono esempi di architetture multiprocessore a ricerca indiretta. La caratteristica qualificante di un approccio di questo tipo è l'elaborazione parallela dei dati su un dispositivo di memorizzazione intermedia.

STARAN [2], [4], pur essendo stato progettato originariamente per l'elaborazione d'immagini, grazie alla sua struttura associativa, illustrata in fig. 8, può essere utilizzato anche per la gestione di basi di dati.

Le possibilità di collegamento con il computer ospite sono molteplici: in particolare, può essere utilizzato un indirizzo di memoria specifico in modo che qualsiasi cosa scritta in quella locazione da un sistema, sia esso l'ospite o STARAN, venga trasmessa all'altro sistema (*External Function Logic*).

Tuttavia, il solo collegamento ad alta velocità, e perciò estremamente più costoso degli altri, è costituito dalla linea parallela di I/O.

La necessità di tale linea è giustificata dell'osservazione che ognuna delle 32 matrici associative disponibili contiene 8 Kbyte. Il *sequential controller* è un microprocessore PDP-11 destinato al controllo delle operazioni interne del sistema e alla gestione della comunicazione con il computer ospite; l'*associative processor controller* è, invece, un componente microprogrammato che comanda le 32 memorie associative, ciascuna costituita da 256 parole di 256 bit. Ad esse corrispondono altrettante coppie di parole di uguale lunghezza, i cui bit rappresentano rispettivamente flag di abilitazione e di indicazione del risultato di un'operazione di ricerca. Ciò permette di limitare ed un insieme definito di parole l'esecuzione di una determinata istruzione.

Tutte le memorie condividono invece due registri, rispettivamente di mascheramento e di confronto; ciò consente di eseguire un'operazione di confronto parallelamente su tutte le matrici associative. È interessante valutare come STARAN risponda ad una richiesta del tipo "elencare tutti i dirigenti del reparto di produzione". Si assuma che tanto il codice relativo alla qualifica quanto il codice di reparto siano contenuti negli stessi 32 byte (256 bit) della registrazione "impiegato". La relazione considerata, o una sua parte, viene caricata nelle memorie associative ed i bit della parola di selezione abilitati in corrispondenza delle tuple appropriate. Nel registro di mascheramento vengono posti a 1 i bit corrispondenti ai campi relativi alla qualifica e al reparto, mentre nel registro di confronto vengono caricati il codice corrispondente alla qualifica di dirigente e quello relativo al reparto di produzione. A questo punto la richiesta considerata, pur riguardando la combinazione di due predicati, può essere soddisfatta con un unico confronto. Se la parte necessaria della registrazione "impiegato" è più lunga di 32 byte, bisognerà utilizzare più parole per una singola tupla riducendo conseguentemente le prestazioni del sistema; tuttavia, i campi su cui vengono eseguite le operazioni devono essere contenuti nella medesima parola di 256 bit. Ciò rappresenta senza dubbio una grave limitazione.

La particolarità di RAP (Relational Associative Processor) [2], [3], [4] è invece quella di utilizzare il cosiddetto "approccio cellulare"; esso consta, infatti, di un certo numero di unità identiche, dette celle, ciascuna operante su una parte della base di dati e collegata alle due che le sono immediatamente vicine. Il loro coordinamento è affidato a una *Set Function Unit* mentre, come già visto in architetture precedenti, è presente un *controller* dedicato alla comunicazione con il computer ospite. L'architettura del sistema può essere raffigurata come in fig. 9.

Ogni cellula è costituita da due processori, una ALU e una *Information Search and Manipulation Unit,* un buffer di capacità limitata ed una traccia di un dispositivo di memorizzazione rotante (tamburo, memorie ad accoppiamento di carica o a bolle); quest'ultimo particolare architetturale indica che RAP, pur rappresentando un capitolo fondamentale nello sviluppo delle *database machine,* è un progetto ormai fortemente datato.

La cellula rappresenta, in sostanza, una unità indipendente, in grado di eseguire un certo insieme di operazioni relazionali; inoltre, dal momento che essa contiene un proprio registro di confronto, è possibile eseguire, diversamente da quanto visto per STARAN, differenti operazioni parallelamente. Il funzionamento di RAP è abbastanza semplice: una certa relazione viene individuata sulle memorie di massa e trasferita sul supporto di memorizzazione rotante e da qui, in corrispondenza ad ogni rotazione, ai *buffer* delle cellule opportune. Se un record corrisponde al criterio di ricerca, vengono settati i bit opportuni nel blocco dei dati e il record viene ricopiato; durante una rotazione successiva, tutti i record marcati possono essere individuati e inviati all'host.

Un progresso rispetto alle macchine fin qui considerate è rappresentato da DBC (*Data Base Computer*) [1], [2], che si distacca profondamente per filosofia architetturale e per prestazioni. Pur non richiedendo significative migliorie tecnologiche, esso costituisce il primo esempio di *hardware-backend.* Il suo principale obiettivo è quello di supportare basi di dati gerarchiche, reticolari o relazionali di dimensione compresa fra i 100 Megabyte e i 100 Gigabyte, utilizzando comuni dischi magnetici; bisogna sottolineare come una delle maggiori limitazioni degli approcci fin qui considerati fosse il costo e, di conseguenza, la limitata capacità delle memorie utilizzate.

DBC presenta due forme diverse di parallelismo: una prima determinata dalla possibilità di elaborare parallelamente le informazioni di un cilindro, grazie alla presenza di un microprocessore per ogni traccia contenuta nel cilindro stesso, una seconda dovuta alla sua architettura di tipo pipeline. DBC, come si può vedere nella figura seguente (fig. 10), è costituito da sette elementi princi-

Fig. 9 - Architettura RAP.

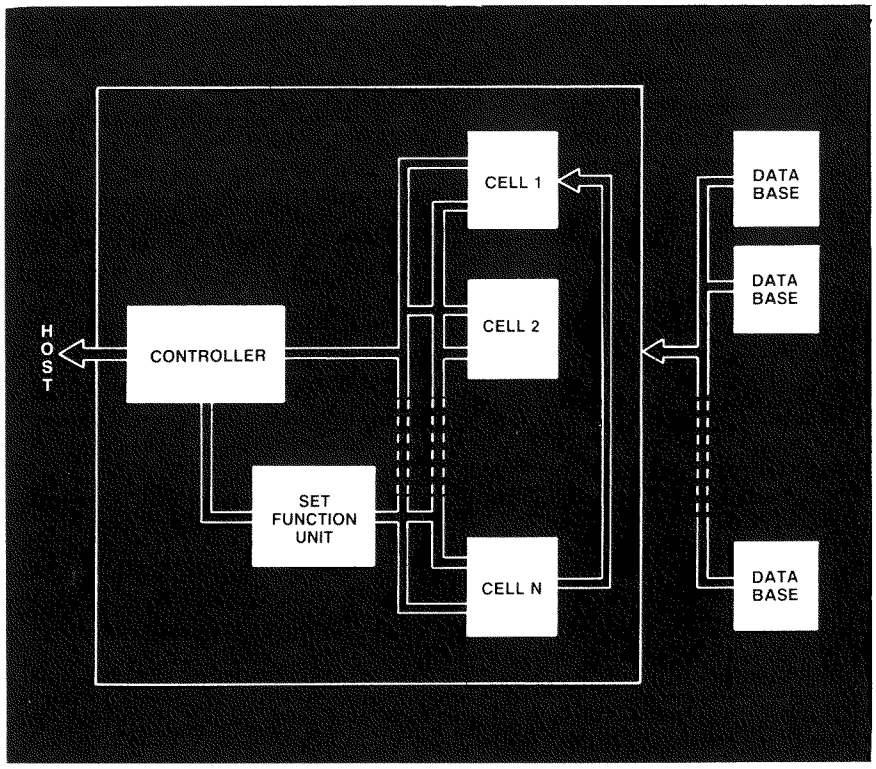
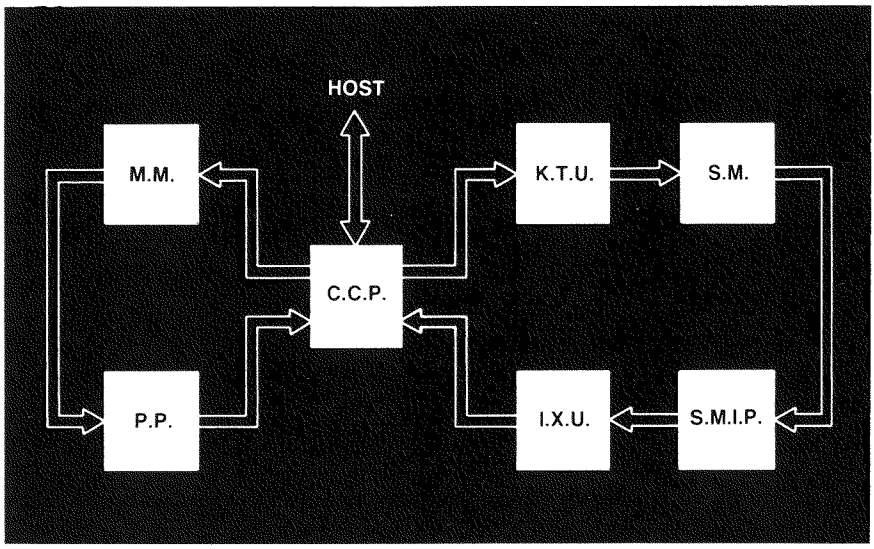


Fig. 10 - Architettura DBC.
Legenda:
C.C.P.: Command and Control Processor
M.M.: database store
P.P.: Post-Processor
K.T.U.: Keyword Tranformation Unit
S.M.: Structure Memory
S.M.I.P.: Structure Memory Information Processor
I.T.U.: Index Translation Unit



pali, divisi in due cicli collegati e controllati da un processore centrale.

Il C.C.P. riceve la richiesta dell'utente dal computer ospite e la converte in una o più operazioni di DBC; quindi, controlla gli altri elementi mentre vengono utilizzati per elaborare la richiesta. Infine esso trasmette i dati richiesti o la conferma di un aggiornamento all'host.

Il ciclo di destra (*structure loop*) ha il compito di trasformare le espressioni di selezione, contenute in una transazione, negli indirizzi fisici dei blocchi di dati da reperire sulle memorie di massa; questi blocchi corrispondono ai cilindri dei dischi magnetici.

La *Keyword Transformation Unit* converte una determinata richiesta in una serie di indirizzi di S.M. che rappresenta, di fatto, una tabella d'inversione per la base di dati; quest'ultima produce gli indirizzi logici dei dati da utilizzare per soddisfare la richiesta. Su tali indirizzi, S.M.I.P. esegue operazioni booleane; l'*Index Translation Unit,* infine, traduce gli indirizzi logici in indirizzi fisici. A questo punto, i dati desiderati possono essere prelevati dal C.C.P. sui supporti di memorizzazione.

Il compito del ciclo di sinistra (*data loop*), e in particolare del *Post Processor,* è quello di garantire una sicurezza maggiore di quella fornita da S.M. che controlla gli accessi a livello di blocco di dati. Dal momento

che il *Post Processor* esamina le registrazioni selezionate, i controlli possono venire eseguiti a livello del singolo dato e del suo valore. Il *Post Processor* inoltre, converte il flusso parallelo di record ricevuti (potenzialmente uno da ogni testina del disco) in un flusso seriale, ordinato nel modo specificato dall'utente.

5. Sviluppi architetturali

DBC ha segnato, senza dubbio, una svolta architetturale nel campo delle *database machine,* generando numerosi progetti che ne rappresentano una diretta estensione; la loro caratteristica comune è quella di eliminare la principale lacuna di DBC: l'impossibilità di eseguire operazioni di tipo joint in modo efficiente. Le soluzioni adottate al fine di risolvere questo problema, possono essere schematicamente suddivise in due gruppi: quelle che prevedono un'interconnessione fra i microprocessori "di traccia", volta a favorire il trasferimento di dati necessario per eseguire un'operazione di *cross reference* su due relazioni, e quelle che, al contrario, postulano l'introduzione di un *join processor* progettato "ad hoc".

La ricerca di nuove soluzioni, che garantiscano prestazioni sempre migliori, non poteva naturalmente limitarsi a produrre filiazioni dirette di DBC. In questo senso vanno progetti come l'italiano DBMAC [9], ad esempio, che presentano architetture innovative e che, tipicamente, accoppiano i vantaggi dell'approccio cellulare (RAP), cioè della replicazione delle coppie memoria-processore, con quelli dell'approccio di specializzazione funzionale, di cui DBC costituisce l'esempio principale. Ciò consente di ottenere una doppia modularità, sia in termini di capacità di elaborazione e memorizzazione che di funzioni eseguibili dal sistema. Una caratteristica di questo tipo è il risultato di un'interessante separazione, durante la prima fase della progettazione, dell'architettura logica da quella fisica. Per quanto riguarda la prima, la specializzazione funzionale permette di definire moduli funzionali a cui sono demandati compiti ben definiti; per quanto riguarda la seconda, la replicazione funzionale consente di definire la struttura multiprocessore del sistema.

Il problema di unificare queste due architetture è risolto per mezzo di un sistema operativo distribuito, attraverso il quale i processi di sistema, definiti nell'architettura logica e localizzati fisicamente su celle diverse, possono comunicare e sincronizzarsi. Non deve sfuggire il fatto che un controllo distribuito di questo tipo, pur nella sua maggiore complessità, sostituisce il processore di controllo tipico delle architetture cellulari; l'eliminazione del *controller*, il cui mancato funzionamento compromette l'intero sistema che lo contiene, comporta una maggiore affidabilità e garantisce, almeno in via teorica, migliori prestazioni.

In definitiva, DBMAC rappresenta compiutamente l'ultima generazione di *database machine,* ormai in grado di eseguire un insieme completo di funzioni per la gestione di una base di dati e di rappresentare un nodo specializzato in una rete di calcolatori. Oltre a ciò, DBMAC garantisce la riservatezza delle informazioni conservando un elenco degli utenti del sistema e dei loro diritti d'accesso tanto in termini di dati quanto di transazioni consentite; nel caso di interazioni ripetitive, esso prevede la possibilità di costruire una libreria di transazioni precompilate. Infine, DBMAC consente elaborazioni parallele a livello di utente, di transazione e perfino di operazione primitiva della base di dati.

Un'altra caratteristica fondamentale di DBMAC è quella di eseguire le operazioni relazionali non sull'intero database, come facevano tutte le macchine sin qui considerate, bensì sugli indici; ciò è possibile grazie all'utilizzazione di una struttura dati opportuna, detta *data pool,* che consente di memorizzare la base dei dati come l'insieme dei domini su cui sono definiti tutti gli attributi. Le relazioni vengono rappresentate sotto forma di informazioni di controllo. Gli insiemi di dati, ciascuno contenente i valori di un dominio e le relative informazioni di controllo, costituiscono il *data pool.* La fig. 11 riporta l'esempio di una base di dati, costituita da tre relazioni e quattro domini, realizzata in questo modo: le informazioni di controllo indicano, per ogni valore del dominio, la relazione e la tupla ai quali esso appartiene.

Una base di dati così organizzata occupa meno spazio rispetto ad una in cui le relazioni sono implementate per mezzo di archivi piatti, le cui registrazioni, cioè, siano mantenute in prima forma normale, come solitamente avviene. I valori del dominio e le informazioni di controllo vengono memorizzate nel *data pool* per mezzo di un albero (*albero del dominio*) di altezza pari a quattro; la radice contiene il nome del dominio, i nodi del primo livello contengono valori del dominio, quelli del secondo indentificatori di relazioni, quelli del terzo infine, identificatori di tuple. In questo modo, ogni percorso radice-foglia identifica un'istanza di un attributo. Questa struttura consente una totale partizione delle relazioni in base agli attributi, incrementando così la capacità di elaborazione parallela del sistema; inoltre, la suddivisione dei domini favorisce l'esecuzione di operazioni di *join.*

DBMAC rappresenta quindi un esempio di quella compenetrazione fra soluzioni hardware e software cui si accennava in precedenza e che è uno degli aspetti salienti di recenti ricerche (si veda, ad esempio, [10]).

6. Vantaggi e svantaggi delle database machine

I possibili vantaggi offerti da una *database machine* appaiono ormai evidenti; il principale è rappresentato, senza dubbio, da un significativo miglioramento delle prestazioni ottenuto, come visto, utilizzando hardware dedicato, elaborazione parallela, memoria associativa ed altre tecniche applicate alla gestione della base di dati.

In generale, le prestazioni vengono migliorate anche riducendo il volume di dati che devono essere trasferiti dai supporti di memorizzazione al computer ospite. Ciò è reso possi-

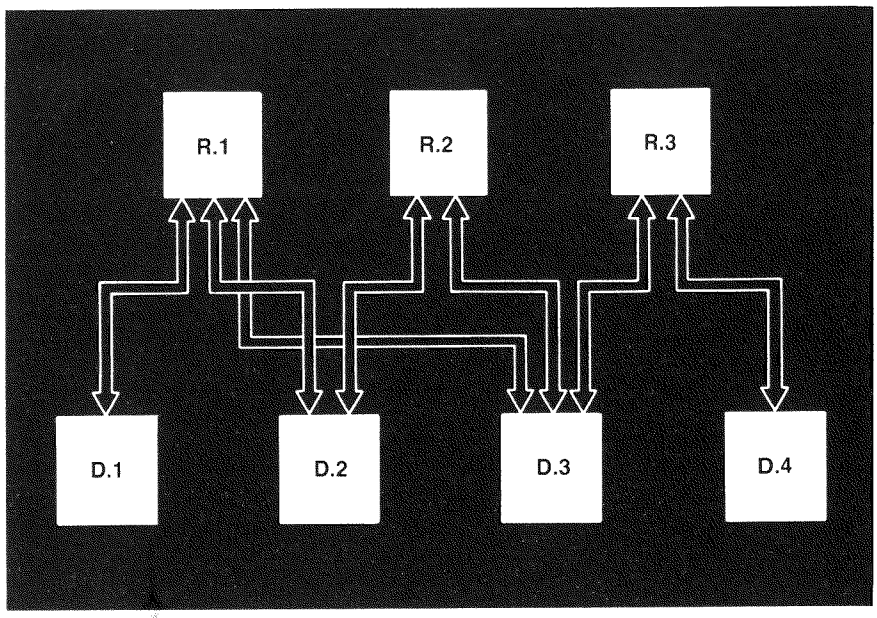


Fig. 11 - Un esempio di base di dati, costituita da tre relazioni e quattro domini:
R.1: (a1 : D.1; a2 : D.2; a3 : D.3)
R.2 : (a4 : D.2; a5 : D.3)
R.3 : (a6 : D.3; a7 : D.4)

bile grazie ad operazioni di *select, project,* compressione o codifica dei dati. Bisogna notare inoltre che la presenza di una *database machine* restituisce al computer ospite una quantità, spesso considerevole, di memoria e tempo di elaborazione, utilizzabile per programmi applicativi; questi ultimi, potendosi interfacciare con la base di dati "ad alto livello", risultano notevolmente semplificati. Tale beneficio è di importanza fondamentale qualora il computer ospite sia prossimo alla saturazione. Inoltre, dal momento che la maggior parte delle interruzioni di I/O e la loro relativa gestione avviene nel *backend*, l'overhead che esse determinano nell'host viene fortemente ridotto.

Infine, una *database machine* costa sostanzialmente meno delle risorse del computer ospite che sostituisce e spesso ne ritarda l'obsolescenza; infatti, essa può eliminare colli di bottiglia dovuti tanto alla velocità del processore quanto alla dimensione della memoria centrale. I benefici funzionali non sono certamente meno importanti: la difficoltà di rendere una base di dati accettabile a più computer può essere, in gran parte, superata con l'introduzione di una *database machine,* poichè è sufficiente fornire un'interfaccia tra quest'ultima ed ogni nuovo host. Ciò è senz'altro più semplice di eventuali modifiche all'intero sistema per la gestione di basi di dati. La maggiore sicurezza e integrità della base di dati sono anch'essi vantaggi facilmente intuibili, poichè una *database machine* isola i programmi applicativi, e più in generale il computer ospite, dalla base di dati. Infine, anche il ripristino del sistema dopo eventuali malfunzionamenti può essere migliorato poichè le due macchine, la *database machine* e l'ospite, possono controllarsi reciprocamente se vengono previsti due *log* separati. In uno, l'ospite memorizza tutte le transazioni e le richieste inviate alla *database machine;* quindi, se quest'ultimo cade, le operazioni di *recovery* possono essere eseguite utilizzando questo *log.* Analogamente, la *database machine* conserva nell'altro tutte le variazioni apportate alla base di dati; se l'host cade, il ripristino del sistema viene eseguito per mezzo di questo secondo *log.*

Questi vantaggi non vanno tuttavia accettati in senso assoluto, a cominciare proprio dal conclamato miglioramento delle prestazioni; perchè esso sia effettivo, infatti, è necessario che almeno la metà delle risorse del sistema siano impiegate nella gestione della base di dati e che ogni operazione comporti, generalmente, un elevato numero di accessi sulle memorie di massa. Se questi vincoli sono rispettati, l'incremento delle prestazioni ottenibile va da un minimo di cinque a un massimo di cinquanta volte [4]. Un'altra osservazione, che può risolversi in un grave problema, è la necessità che la *database machine* sia in grado di continuare a funzionare, con una non eccessiva riduzione delle proprie prestazioni, in presenza di qualche guasto; infatti gran parte dei vantaggi che ne giustificano l'introduzione vengono meno se essa non costituisce l'unica via d'accesso ai dati.

Un'ultima obiezione che può essere sollevata, riguarda il non facile dimensionamento del sistema; le sue risorse, infatti, non possono essere liberamente distribuite, poichè una sua parte, nella fattispecie la *database machine,* è dedicata. Le prospettive di un aumento del carico del sistema devono essere perciò valutate con attenzione per evitare dannosi sbilanciamenti fra computer ospite e *database machine*; la saturazione di una delle due macchine produrrebbe un sottoutilizzo dell'altra.

Pur con le cautele che queste ultime osservazioni suggeriscono, esistono due fondamentali fattori che sembrano destinati a favorire la diffusione e lo sviluppo delle *database machine:* da un lato il crescente successo delle reti di calcolatori con la conseguente necessità di fornire grandi

basi di dati distribuite, condivisibili da tutti i computer collegati, dall'altro l'esigenza di migliorare le ancora spesso insufficienti prestazioni delle basi di dati relazionali. Le *database machine* si propongono come un'efficace soluzione ad entrambi questi problemi. In particolare, per quel che riguarda il primo, diversamente da quanto avviene in un approccio convenzionale in cui grosse porzioni di archivi vengono trasferiti al richiedente, quando una *database machine* è utilizzata come sottosistema di memorizzazione, i dati vengono acceduti e selezionati nel sottosistema stesso. Inoltre, per permettere la condivisione degli archivi è sufficiente utilizzare un singolo insieme di semafori; la gestione di questi ultimi, difficile in un sottosistema di memorizzazione non intelligente, risulta semplificata dalla presenza delle risorse di calcolo proprie di una *database machine.*

Per quanto riguarda il futuro, le prestazioni delle *database machine* sono destinate a crescere di pari passo con i progressi della tecnologia hardware; ciò potrebbe rivelarsi un vantaggio decisivo nella competizione che le vede opposte ai sistemi di gestione di basi di dati tradizionali ed al tempo stesso costituire una spinta sufficiente a vincere la resistenza che gli utilizzatori spesso oppongono a proposte fortemente innovative come queste.

7. Bibliografia

[1] J. BANNERJE, D.K. HSIAO e K. KANNAN, *"DBC-A database computer for very large databases"*, IEEE Trans. on Computers, pp. 414-429, Vol. C-28, June 1979.

[2] O.H. BRAY, H.A. FREEMAN, *"Data Base Computers"*, Heath Company, 1979.

[3] G.A. CHAMPINE, *"Four Approaches to a Data Base Computer"*, Datamation, pp. 101-106, Dec. 1978.

[4] G.A. CHAMPINE, *"Current Trends in Data Base Systems"*, IEEE Computer, pp. 27-41, May 1979.

[5] G.A. CHAMPINE, *"Back-End Technology Trends"*, IEEE Computer, pp. 50-58, February 1980.

[6] R. EPSTEIN, *"Why Database Machines?"*, Datamation, pp. 139-144, July 1983.

[7] H. HE, M. HIGASHIDA e altri, *"The Implementation of a Multibacked Database System (MDBS)"*, pp. 327-385, in [8].

[8] D.K. HSIAO, *"Advanced Database Machine Architecture"*, Prentice-Hall, 1983.

[9] M. MISSIKOFF, M. TERRANOVA, *"The Architecture of a Relational Database Computer known as DBMAC"*, pp. 87-107, in [8].

[10] S. PRAMANIK, F. FOTOUCHI, *"Index Database Machine"*, The Computer Journal, pp. 451-457, Vol. 29, No. 5, 1986.

[11] M. VANNESCHI, *"Architettura dei Sistemi di Elaborazione"*, Vol. 1, Servizio editoriale universitario di Pisa, 1983.

Il software didattico nella formazione scolastica e professionale

MICHELE PELLEREY

Università Pontificia Salesiana
Facoltà di Scienze dell'Educazione
Roma

1. La prima ondata: l'interesse per il nuovo oggetto tecnologico

La possibilità di usare il computer per favorire l'apprendimento sia scolastico, sia professionale, fu percepita abbastanza presto. È facile constatare che si usarono i paradigmi di riferimento psicopedagogici allora più in voga nel campo della ricerca, soprattutto in ambito universitario e militare. Erano i tempi dell'Istruzione Programmata lineare Skinneriana e di quella ramificata Crowderiana. Per inciso: quest'ultima si sviluppò per rendere rapida ed efficace la preparazione dei riparatori di radar durante la guerra. Lo stesso progetto Plato e il linguaggio Basic per l'uso dei computer in timesharing nacquero secondo una filosofia formativa specifica e con un pregiudizio: poter sostituire il docente o il formatore.

La diversificazione degli oggetti tecnologici didattici (hardware e software) con l'avvento dei personal, se da una parte sembrò aprire una pluralità di prospettive di uso pedagogico didattico del computer, dall'altra restrinse l'attenzione alla descrizione e valutazione dell'oggetto in sé. Per questo Papert parlò di tecnocentrismo nell'analisi dei risultati dell'uso del software nella formazione [4]. D'altro canto spesso l'oggetto tecnologico porta in sè inscritta una impostazione pedagogico-didattica. Nel caso di Papert si ha un esempio significativo di stretta connessione tra sviluppo del software e sollecitazione di una linea pedagogico-didattica precisa: infatti, egli può essere collocato nel movimento piagetiano riletto sotto la prospettiva della didattica della scoperta.

La mia relazione tende a rovesciare la prospettiva e, anzichè partire dall'oggetto tecnologico e dalla filosofia pedagogico-didattica in esso inscritta o a esso strettamente legata, prende le mosse dalla cultura pedagogico-didattica che è presente nelle nostre azioni formative e analizza, quindi, ove e come si collocano i diversi tipi di software in relazione a essa. Per concludere un po' polemicamente con un appello al pluralismo e alla relativizzazione del metodo da un punto di vista ideologico, per riportarlo ai suoi veri compiti: facilitare l'apprendimento e favorire sia l'efficienza, sia l'efficacia dell'azione formativa.

Per far questo presenterò tre diagrammi di riferimento, uno a carattere più psicologico, l'altro più pedagogico, il terzo più direttamente riferito ai processi di formazione professionale.

2. Un primo strumento di analisi: D.P. Ausubel (1968)

Un primo strumento di analisi del processo di acquisizione delle conoscenze e delle competenze sia scolastiche che professionali è stato elaborato negli anni sessanta da D.P. Ausubel [1], un noto studioso di psicologia dell'educazione.

È uno schema che deriva da una precisa teoria psicologica elaborata sulla base dei contributi di autori come M. Wertheimer, G. Katona e suoi personali: teoria polemica con l'allora imperante pedagogia della scoperta.

Nello schema di analisi vengono distinte due fondamentali dimensioni relative ai processi di acquisizione della conoscenza e della competenza.

La prima dimensione riguarda il modo secondo il quale la nuova conoscenza, o la nuova abilità, viene incorporata nell'insieme dei fatti, dei concetti e delle generalizzazioni già interiorizzati e organizzati dall'allievo. Ai due estremi di questa dimensione stanno un apprendimento meccanico e uno significativo.

La seconda dimensione concerne l'itinerario attraverso il quale le nuove conoscenze e abilità vengono acquisite. Ai due estremi di questa dimensione stanno l'apprendimento per scoperta e quello per ricezione. Oggi si parla meno di apprendimento per scoperta, l'equivalente che possiamo usare è apprendimento per ricerca: ricerca d'altronde più o meno autonoma.

a) Un apprendimento meccanico è caratterizzato dal fatto che il nuovo elemento di conoscenza viene acquisito in maniera isolata, senza connessioni o legami con quanto già si conosce. A causa di questa mancanza di relazione tra il nuovo apporto e la struttura conoscitiva già sviluppata, la sola maniera praticabile per poterlo ritenere nella memoria consiste nella ripetizione meccanica

e stereotipata. Non si ha alcuna trasformazione nè dell'apporto conoscitivo esterno, nè della struttura conoscitiva interna. Possiamo aggiungere che, rimanendo la nuova conoscenza isolata nell'insieme dei concetti posseduti, essa risulta difficilmente reperibile e utilizzabile, a meno di tecniche mnemoniche assai artificiali.

b) L'apprendimento significativo, invece, è caratterizzato dal fatto che il materiale da apprendere può collegarsi, e viene di fatto collegato, con gli altri concetti e le altre capacità già posseduti e quindi incorporato non in maniera isolata, ma in forma ben connessa con la struttura conoscitiva precedente. Sia questa che il nuovo apporto vengono più o meno trasformati. La rete di relazioni che così si costituisce rende facile sia il ricordo che l'utilizzazione della nuova conoscenza.

c) L'apprendimento per scoperta (o per ricerca autonoma) mette invece l'accento sull'itinerario di apprendimento e più precisamente sul suo carattere di processo indipendente e sul fatto che la conquista della verità di una proposizione avviene per via autonoma. Non si tratta, nella normalità dei casi, di trovare qualcosa per primi, cioè in senso assoluto, bensì in senso relativo e cioè come se si fosse i primi a farlo. Ciò significa che la soluzione del problema, la dimostrazione della verità di un'affermazione, o un procedimento di lavorazione non ci sono state comunicate da nessuno. Va però precisato che un'autentica scoperta, anche in senso relativo, richiede non solo l'enunciazione di una ipotesi attendibile, ma anche l'effettiva verifica del suo valore di verità. Qui si dovrebbe aprire una serie di considerazioni di natura epistemologica, ma il discorso ci porterebbe lontano.

d) L'apprendimento per ricezione viene di conseguenza identificato in ragione della mancanza di autonomia nella conquista di un concetto, di un procedimento risolutivo, della verità o meno di una proposizione. Tutto questo infatti deriva dalla trasmissione culturale, o comunicazione diretta.

Le asserzioni agli estremi delle due dimensioni non devono essere considerate come dicotomiche. Infatti, le due coppie: "Apprendimento significativo - Apprendimento meccanico" e "Apprendimento per ricezione - Apprendimento per scoperta" non si escludono a vicenda, bensì vanno considerate come poli estremi di una gamma continua di differenti tipi di apprendimento. Si passa, infatti, gradualmente dall'uno all'altro. La polemica, a cui ho accennato, nei riguardi della pedagogia della scoperta sta nell'aver dimostrato da parte di D. P. Ausubel che è possibile un apprendimento significativo per ricezione e d'altra parte che è fattibile un apprendimento meccanico per scoperta, cosa quest'ultima evidente nel caso della ricerca per tentativi ed errori, alla cieca. Ciò basta ad affermare che le due dimensioni sono indipendenti tra di loro e quindi possono combinarsi dando luogo a un diagramma del tipo rappresentato in fig. 1.

In questo diagramma esistono quattro quadranti ed è possibile indicare su di essi il punto che rappresenta il tipo di apprendimento che intendiamo promuovere o di fatto viene realizzato dagli allievi. Questi quattro quadranti servono anche per analizzare il software didattico oggetto di valutazione. Cioè se esso è prevalentemente orientato a un apprendimento per ricezione significativa, come è spesso il caso di alcuni tutoriali ben organizzati; oppure tende a sviluppare la capacità di ricerca e soluzione di problemi e quindi iscrivibile nel quadrante in alto a destra, come è il caso di alcuni software di problem solving, di problem posing o di decision making. Tuttavia è possibile che questi stessi software, in sè collocabili nei quadranti in alto, una volta mal utilizzati o posti al di là delle capacità degli allievi, si trasformino in giochi senza senso e pure esplorazioni a caso e quindi vadano collocati nel quadrante in basso a destra.

Molti vedono di mal occhio software didattici che debbono essere collocati nel quadrante in basso a sinistra e che quindi si ispirino a tecniche ripetitive o ispirate all'Istruzione Programmata. Tuttavia nelle attività sia scolastica, ma soprattutto professionale, occorre molto spesso sviluppare abilità, sia intellettuali che pratiche, che devono essere interiorizzate fino a diventare automatiche e meccanicamente ripetibili per permettere attività di livello superiore.

D'altro canto non tutto il software, come tale, è collocabile in uno solo dei quadranti e per esso può essere identificato un punto preciso di collocazione. Molto dipende dal ruolo che esso deve svolgere nel progetto formativo e dal modo con cui viene utilizzato dall'insegnante o dal formatore.

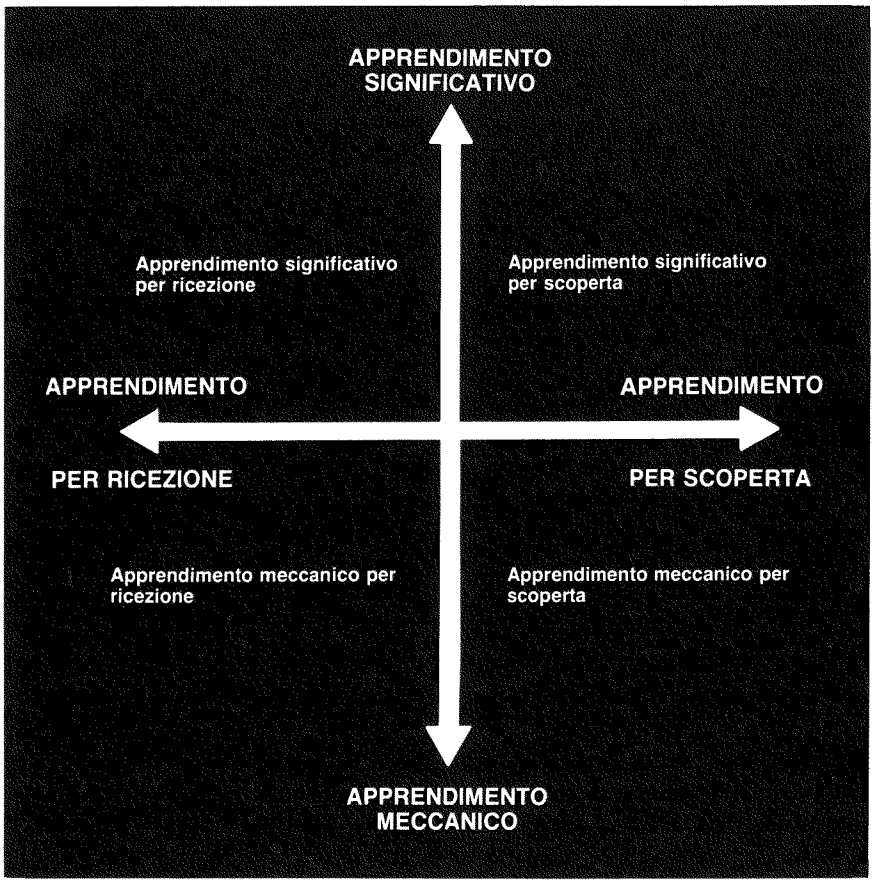
4. Un secondo strumento di analisi: M. Schiro (1978)

Il secondo diagramma è stato elaborato da M. Schiro [2] al fine di fornire un quadro di riferimento all'analisi delle premesse o concezioni pedagogiche che stanno alla base delle scelte di impostazione dei processi formativi sia scolastici, sia aziendali. In esso sarà facile riconoscere molte delle tendenze manifestatesi negli ultimi decenni. Anche in questo caso il diagramma servirà nel nostro caso come guida all'analisi degli eventuali vincoli di natura pedagogico-didattica inscritti nel software didattico, oppure delle possibilità di un suo uso, quando questo si presenti più aperto e flessibile.

Il diagramma è basato, come il precedente, su due dimensioni fondamentali. La prima di esse pone ai suoi estremi la realtà che viene privilegiata nel processo formativo: la realtà oggettiva, esterna, data, sia essa costituita dal mondo culturale codificato nelle discipline oppure dal mondo fisico e sociale, oppure la realtà soggettiva, sia essa quella dell'allievo oppure dell'insegnante o dell'intero gruppo docente.

La seconda dimensione ha ai suoi estremi invece la fonte della conoscenza che si vuole privilegiare oppure il suo uso. A una estremità si colloca quindi l'origine dei concetti,

Fig. 1 - Lo schema di Ausubel.



dei principi, delle teorie o dei procedimenti presi in considerazione; mentre all'altra estremità viene posto l'uso che di questi si fa o almeno si vorrebbe o si dovrebbe fare.

La composizione delle due dimensioni fornisce un diagramma del tipo illustrato in fig. 2.

Sull'asse verticale si passa dalla valorizzazione della fonte della conoscenza a quella del suo uso; sull'asse orizzontale dall'accettazione, come base del processo formativo, della realtà oggettiva fino alla prevalente considerazione di quella soggettiva.

Esaminiamo ora i quattro quadranti che derivano dalla composizione di queste due dimensioni.

Nel quadrante in alto a destra è possibile collocare un insieme di tendenze e affermazioni che privilegiano l'allievo stesso come origine fondamentale della conoscenza e della competenza e, in genere, come perno quasi esclusivo dell'intero processo didattico. Nel caso scolastico ciò indica l'accettazione incondizionata degli interessi e delle aspirazioni dell'educando. Questi ha già in sé un patrimonio culturale e operativo originario e germinale, che ha solo bisogno di non essere impedito nel suo sviluppo. La sua interazione con l'ambiente interno, ma soprattutto esterno alla scuola è la principale fonte dei suoi apprendimenti. Nel caso aziendale il quadrante implica non solo il fatto che è l'esperienza accumulata dal lavoratore o dal manager la base insostituibile in ogni ulteriore acquisizione professionale, ma anche che egli ne è il principale artefice, spesso in contrasto stesso con le strutture formative aziendali.

In questo quadrante si può collocare il Logo considerato all'interno della filosofia educativa del Papert di Mindstorms [3].

Il quadrante in alto a sinistra comprende le posizioni pedagogiche che insistono sul valore formativo delle discipline o delle conoscenze culturali e tecnologiche organizzate e codificate in modo sistematico. Il sapere e il saper fare da acquisire ha una componente sostanziale o semantica, ma anche una componente sintattica e formale, che indica il modo di codificarne e di giustificarne le varie parti. Il processo formativo consiste essenzialmente nella promozione di un'acquisizione sistematica e stabile di questi contenuti conoscitivi e operativi. In questo quadrante possono essere collocati i software di simulazione relativi a processi tipici delle discipline scolastiche o professionali e tutorali che tendono a promuovere l'acquisizione dei loro concetti e principi fondamentali.

Il quadrante in basso a sinistra include in primo luogo una tendenza tecnologica oggi assai diffusa. Nel campo aziendale esso può evocare uno strumentalismo scientifico, ad esempio, per intenderci, l'organizzazione tayloristica del lavoro. Nel campo scolastico un'impostazione attenta alla costruzione di percorsi ben definiti negli obiettivi, nei contenuti, nei metodi, nei mezzi didattici e nelle forme di controllo e regolazione del processo didattico. L'Istruzione Programmata, Skinneriana o Crowderiana, possono essere qui collocate; così come i primi, e molti degli attuali sviluppi del CAI.

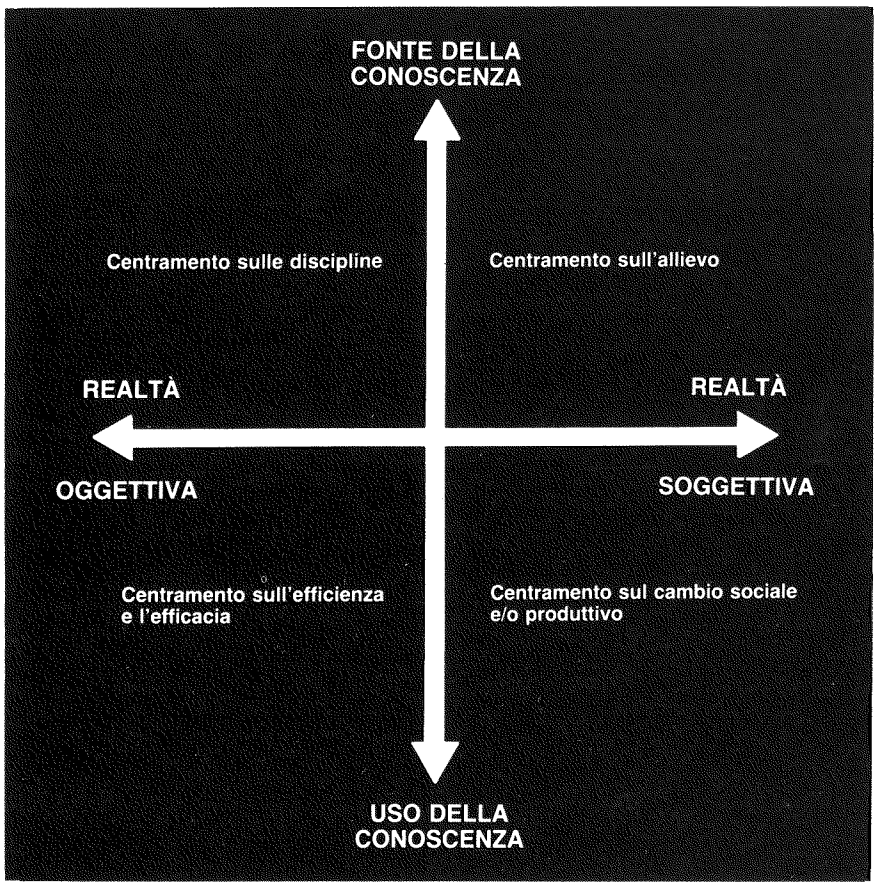


Fig. 2 - Lo schema di Schiro.

Nel campo aziendale si è più attenti all'acquisizione di abilità connesse con le esigenze degli attuali processi produttivi, sotto forma, in questo caso, di addestramento, che all'assunzione di possibili ruoli professionali evolutivi o proiettati sul futuro.

Nel quadrante in basso a destra si possono includere i movimenti che hanno visto la scuola come agente di cambio sociale, se non di attività sovversiva e di rivoluzione culturale. L'impegno educativo scolastico è visto sotto il profilo sociale e in termini di utopia. Si cerca una scuola alternativa, dove i rapporti sociali siano profondamente diversi e ribaltati i processi di comunicazione e di conquista collettiva del sapere. Una versione meno politicizzata valorizza l'interazione e il dialogo come base dello sviluppo della conoscenza e della competenza. In campo aziendale, a parte analoghe posizioni ideologiche e politicizzate, si possono richiamare le esperienze dettate dalla psicologia sociale, dalla dinamica di gruppo, dalle tecniche di facilitazione della comunicazione sia orizzontale, che verticale. Il software prodotto che si ispira a questa filosofia, tende a sviluppare competenze generali di tipo creativo e relative a futuri ruoli professionali o ad ambienti di lavoro modificati in senso ritenuto più soddisfacente per il soggetto.

5. Un terzo strumento di analisi: M. Pellerey (1988)

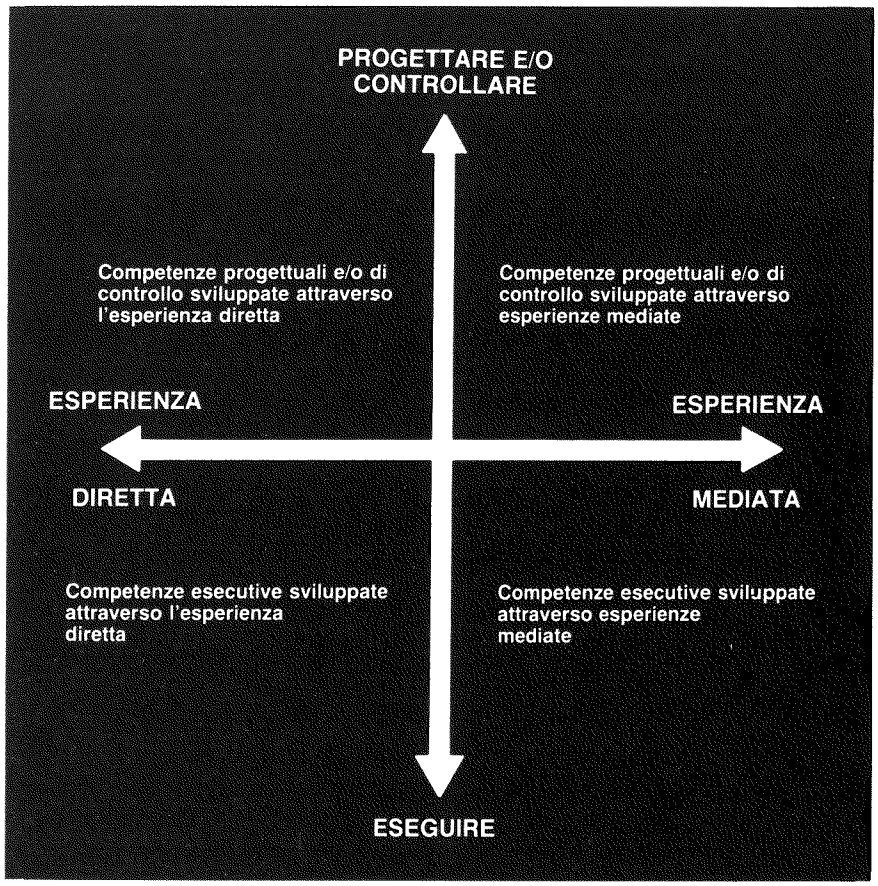
Il terzo strumento di analisi è più direttamente collegabile con le attività tipicamente aziendali. È una proposta che deve ancora essere discussa e messa a punto perchè appena abbozzata e in qualche maniera ancora soggetta a ripensamenti e miglioramenti. Essa si presenta ancora come un diagramma a due dimensioni. La dimensione verticale contrappone ai suoi estremi le finalità che si pongono nelle attività di formazione aziendale, oppure in quelle tecnico-professionali proprie di un corso di istruzione secondaria, e cioè se essere capaci di eseguire una procedura più o meno complessa e coinvolgerne un certo numero di parametri e di realizzare, quindi, un prodotto secondo standard prefissati, oppure possedere una competenza adeguata nel progettare sia un prodotto che un processo di produzione e/o nel controllarne la qualità. Evidentemente nei casi specifici andrebbe evidenziato il settore professionale coinvolto.

La seconda dimensione evidenzia invece la modalità attraverso la quale si desidera, o si è costretti, a realizzare il processo di acquisizione delle competenze professionali in oggetto: se attraverso l'esperienza diretta sul posto di lavoro, molto spesso sotto la guida di personale più esperto; oppure se fuori dall'ambiente specifico di lavoro per mezzo di una molteplicità di interventi e di forme comunicative spesso di natura multimediale e simulativa.

Ritengo inutile insistere sul significato dei quattro quadranti che in questo modo vengono a determinarsi. (fig. 3).

Anche in relazione a questo diagramma è possibile analizzare i diversi software, ma soprattutto il diverso uso formativo che di essi si può fare.

Fig. 3 - Lo schema di Pellerey.



6. Qualche conclusione e conseguenza di un discorso appena avviato

La ricerca di un metodo didattico universale, tale cioè da garantire la sua efficacia di fronte a qualsiasi docente e a qualsiasi allievo è stato un mito inseguito dal tempo di Comenio.

Herbart stesso, che vedeva la matematica come una scienza paradigmatica, da questa traeva lo spunto per individuare un metodo didattico che garantisse esso stesso, se applicato correttamente, i risultati didattici. È il metodologismo che si è sviluppato non solo in matematica, ma anche nelle scienze della natura e via via di là esteso a molti altri ambiti del sapere.

Trovare l'efficacia e l'efficienza nel metodo, anzi in un particolare metodo universale. Si potrebbero ripercorrere i diagrammi sopra riportati per indicare i corifei di questa impostazione, che via via si sono succeduti. Anche in questo caso occorre rovesciare i termini del problema. Non è il metodo che garantisce l'efficacia e l'efficienza dei risultati, bensì i risultati effettivamente ottenuti, e i relativi costi ragionevoli in termini di risorse economiche, strumentali e umane, che indicano la bontà o meno del metodo adottato.

Anche nel campo della didattica che utilizza il computer e il relativo software per realizzare o migliorare i processi formativi, occorre porsi sotto questa prospettiva. Non esiste un software didattico in sè migliore di un altro per motivi teorici o ideali. Anche un software mediocre può ottenere buoni risultati in mano a formatori validi.

È l'insieme del sistema formativo, composto da persone, ambienti, tempi, strumenti di vario tipo, ma soprattutto dagli allievi con il loro patrimonio conoscitivo e di competenze, le loro motivazioni e aspirazioni, che determinano i risultati di un processo formativo.

Celebrare un software didattico, a qualunque schiera tra quelle indicate appartenga, come capace di produrre risultati automatici a prova di utente e di docente, somiglia a certa propaganda che promette di far apprendere una lingua straniera "dormendo".

Quanto affermato non intende però misconoscere il ruolo e il valore di una tecnologia di produzione del software didattico, tecnologia certamente intesa in senso vasto e pregnante. Tecniche di progettazione, di implementazione grafica e interattiva, di valutazione formativa e di qualità, di verifica sul campo, vanno sviluppate e migliorate, tenendo sempre presente però una visione complessiva e comprensiva delle potenzialità e dei limiti di ogni risorsa didattica e formativa.

Non viene quindi contestato il senso di una valutazione del prodotto tecnologico in sè; bensì viene ricordato, come d'altronde la storia dei brevetti e dell'assorbimento di essi nel mercato insegna, che non basta aver realizzato in modo geniale un nuovo prodotto, occorre che esso possa e venga di fatto accolto come valido e funzionale nel contesto mercantile per il quale esso è stato concepito.

D'altro canto è del pari significativa la storia della produzione e adozione dei libri di testo e dei materiali didattici nella scuola e nella formazione.

A quanto detto si può ora aggiungere una postilla. Occorrono sempre più strumenti che consentano di rappresentare, controllare e governare le proprie conoscenze e competenze: come il porsi delle domande, pianificare e guidare la propria attività, correggersi, controllare lo stato delle proprie conoscenze, della comprensione, delle competenze, selezionare strategie valide e produttive nella soluzione di problemi o come risposta a stimoli esterni, ecc. In genere, quando si ha un processo di autoregolazione si pensa che si manifesti una situazione di metacognizione.

Il problema centrale è come si sviluppi la metacognizione, cioè come si acquisisca questa capacità di conoscere, valutare e utilizzare le proprie conoscenze e competenze. La ricerca psicologica ha già dato alcune risposte interessanti. Dal punto di vista della formazione sia scolastica che aziendale questa capacità di autorappresentazione, di autocontrollo e di autogoverno sia della conoscenza sia della competenza è della massima importanza. Credo sia possibile sviluppare software didattico che consenta non solo di sviluppare queste capacità di metacognizione, ma di permetterne l'utilizzazione sistematica nella propria attività d'apprendimento scolastico e professionale. È questa una nuova frontiera per la nostra attenzione e il nostro impegno.

Nota

I contenuti di questo articolo sono stati originariamente presentati dall'Autore a DIDAMATICA 88 (Milano, aprile 1988).

Bibliografia

[1] AUSUBEL D.P. (1968), *Educational Psychology: A Cognitive View,* New York, Holt Rinehart & Winston.

[2] SCHIRO M. (1978), *Curriculum for better schools: the great ideological debate,* Englewood Cliffs, N.J., Educational Technology Publ.

[3] PAPERT S. (1980), *Mindstorms: children, computers and powerful ideas,* New York, Basic Books.

[4] PAPERT S. (1987), *Computer criticism vs. technocentric thinking,* in Information technology and Education, Jan, 22-30.

Sistema esperto per il progetto di reti di comunicazione (*)

M. BEATRICE GOZZOLI,
GIACOMO ZANOTTI

Honeywell Bull Italia
Serv. Supporto Networking
Milano

1. Introduzione

Il Network Design è una disciplina che ha oggi un rilievo di primaria importanza nell'area delle telecomunicazioni: questo deriva dal crescente sviluppo di reti private per la trasmissione dati, geograficamente vaste e applicativamente complesse.

La previsione del comportamento della rete dati è essenziale al fine di garantire all'utente buone prestazioni, salvaguardando gli investimenti, solitamente molto elevati.

Si vuole qui descrivere un'applicazione delle tecniche di intelligenza artificiale rivolte al Network Design: il progetto, in fase di realizzazione alla Honeywell Bull Italia, nasce dalla cooperazione tra il servizio Supporto Networking, che fornisce le specifiche funzionali, e la Direzione Sistemi Esperti, che cura l'implementazione.

Gli aspetti presi in esame in questa presentazione sono rivolti alle problematiche affrontate dal Sistema Esperto e non ai dettagli di tipo implementativo dello stesso.

2. Network Design

Scopo del Network Design è la definizione della rete ottima, quella, cioè, di costo minimo e prestazioni massime.

Lo studio di una rete di telecomunicazioni verte, pertanto, sul disegno topologico, il dimensionamento di linee e nodi, l'analisi dei ritardi e degli instradamenti del traffico, il calcolo di costi fissi e canoni, la valutazione del ritorno degli investimenti.

Tutta questa attività comporta una serie di calcoli estremamente complessi, effettuati tramite procedure di tipo iterativo e modelli matematici basati sulla teoria delle code.

Se da un lato svolgere manualmente questa mole di lavoro consente una maggior aderenza ad alcune realtà delle telecomunicazioni in Italia, dall'altro comporta tempi estremamente lunghi soprattutto nel caso di reti particolarmente complesse.

Viceversa l'automazione delle procedure abbatte in modo drastico i tempi di analisi e sintesi a scapito, però, di una maggiore precisione dei risultati: questo deriva infatti dall'utilizzo di modelli e algoritmi non sempre aderenti alla realtà: alcune parti dell'analisi dipendono strettamente dalla struttura della rete telefonica o della rete dati di un determinato paese.

A titolo di esempio, in Italia, i portanti fisici seguono cammini ben precisi di cui si deve tener conto soprattutto per questioni legate all'affidabilità della rete; il disegno delle linee multipunto deve seguire i criteri gerarchici della rete telefonica (settore, distretto, compartimento).

Va anche aggiunto che la maggior parte dei modelli ipotizzano tecniche di instradamento di tipo proporzionale mentre normalmente i sistemi nodali prevedono strategie di routing di tipo diverso: in particolare i sistemi nodali Datanet-8 della Honeywell Bull utilizzano una tecnica di routing di tipo a coda più breve.

Da quanto sopra esposto, deriva che il processo di analisi di rete, effettuato in modo manuale o automatico, comporta comunque degli inconvenienti per nulla trascurabili.

La condizione ideale sarebbe quella di utilizzare procedure automatiche in cui i risultati, provenienti dai processi algoritmici, possano venir manipolati da strategie decisionali basate sull'esperienza di persone competenti nel campo delle telecomunicazioni.

In questo risiede la scelta di utilizzare le tecnologie dell'intelligenza artificiale per lo sviluppo di software atto allo studio, analisi e dimensionamento di una rete di trasmissione dati.

3. Il sistema esperto

Il sistema esperto WANESSA (Wide Area Network Expert System Assistant) intende fornire una soluzione al problema di analisi e sintesi di una rete dati privata, di tipo geografico e con modalità operativa a commutazione di pacchetto.

L'intero problema, da un punto di vista matematico, può essere così formulato:

A) dati:

- le locazioni di host e terminali
- la matrice del traffico

(*) Presentato alla Giornata di studio AICA-AEI su "Applicazioni industriali dei Sistemi Esperti" (Milano, giugno 1988).

- la matrice dei costi
- la lunghezza media dei pacchetti
- la velocità di elaborazione dei nodi

B) con i vincoli:
- sul ritardo
- sulla affidabilità

C) e sulle variabili:
- topologia
- capacità delle linee
- procedure di instradamento

D) raggiungere l'obiettivo di:
- minimizzare i costi.

In altri termini il goal di WANESSA è la soluzione di quello che in letteratura è indicato come problema TFCA (Topology Flow and Capacity Assignment).

Va anche evidenziato che il progetto WANESSA si limita all'analisi di uno specifico tipo di rete le cui caratteristiche primarie sono:

1 - rete geografica, centralizzata o distribuita;

2 - portanti fisici privati di rete telefonica o numerica in Italia;

3 - modalità operativa X.25;

4 - tecniche di instradamento del traffico di tipo a coda più breve (nell'ambito dei rami di uscita su cui sono aperti circuiti virtuali diretti ad uno specifico destinatario).

Il sistema WANESSA è scritto in linguaggio LISP per un ambiente PC sotto sistema operativo MS-DOS.

Come già accennato, l'utilizzo di tecniche di Intelligenza Artificiale per la risoluzione del problema è dettata principalmente dalla necessità di manipolazione dei dati, provenienti da processi algoritmici, in base a strategie decisionali non ben definibili e, comunque, assai complesse se espresse mediante i tradizionali metodi di programmazione.

Risulta altresì di difficile valutazione il ruolo d'importanza che l'esperto assegna ai vari fattori in gioco nel processo decisionale: molto spesso le scelte che vengono fatte in uno studio dipendono da elementi noti "inconsciamente" all'esperto o che riflettono condizioni variabili da rete a rete, da cliente a cliente, da intenti commerciali diversificati. Il problema è di come valutare questi fattori per inserirli con il dovuto ruolo di importanza nella "base di conoscenza", che è il fulcro del sistema esperto.

Va anche detto che l'utilizzo della programmazione funzionale consente di modificare in maniera relativamente facile il livello di significatività dei diversi parametri a differenza di quanto succede con la programmazione tradizionale in cui la modifica ad un processo decisionale potrebbe comportare notevoli rifacimenti di codice software.

4. Struttura del progetto WANESSA

4.1 Modelli e algoritmi

Nessun modello è attualmente disponibile e in grado di rappresentare l'intero problema TFCA: l'approccio è, pertanto, quello di suddividere il problema originario in sottoproblemi più semplici.

Vanno anzitutto premesse alcune considerazioni:

1 - se anche questi sottoproblemi fossero tutti risolvibili con metodologie esaustive, non sarebbe garantita l'ottimalità globale della soluzione per l'intero problema;

2 - taluni sottoproblemi, ad esempio quello della determinazione della topologia, possono essere risolti con metodi esatti ma con una mole di calcoli irragionevole; ad esempio, una rete di n nodi ammette $2^{n(n-1)/2}$ topologie diverse: se n è 10, vi sono 4 x 10^{13} soluzioni da calcolare ed esaminare;

3 - le metodologie utilizzate per la risoluzione dei sottoproblemi sono di tipo euristico: questo comporta processi iterativi che generano soluzioni locali sempre più vicine alla soluzione ottima. Tale strategia consente di affrontare problemi per cui non esiste un processo esaustivo o in cui tale processo richiede un numero troppo elevato di calcoli;

4 - perturbando alcune variabili dei vari sottoproblemi, si possono generare soluzioni distinte tra cui, scartate quelle dominate, si può scegliere quella più vicina alle richieste del cliente.

La costruzione della rete avviene, quindi, per passi successivi, alla fine di ciascuno dei quali l'utente ha, comunque, la possibilità di modificare alcune condizioni cambiando il valore di variabili in input o forzando soluzioni specifiche.

Più in dettaglio il processo WANESSA si sviluppa nei seguenti passi:

step 1 determinazione dei punti di concentrazione della rete; si utilizza l'algoritmo ADD [1;2]. La soluzione ottenuta può essere perturbata modificando il parametro in input, "candidatura a centro di concentrazione", assegnato ad ogni località o l'output stesso del processo;

step 2 determinazione della topologia iniziale sulla base del livello di connettività [2];

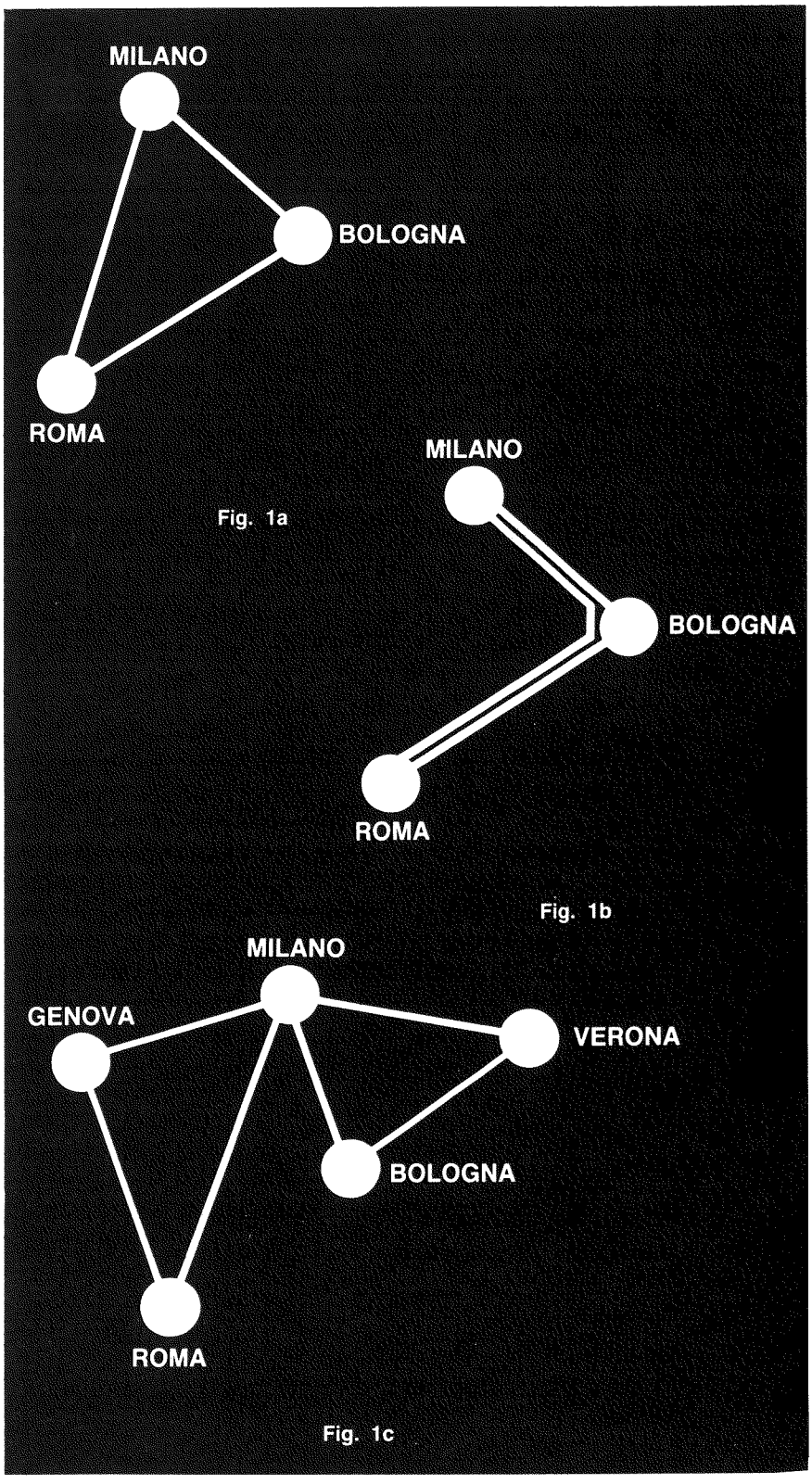
step 3 assegnazione di flussi e capacità iniziali; si usa l'algoritmo Flow Deviation [3], iterando il processo, in base ad una predefinita tolleranza fino ad una soluzione ottima;

step 4 determinazione della topologia finale mediante l'algoritmo Cut Saturation [4]; le perturbazioni sulla soluzione avvengono tramite l'algoritmo stesso o interventi forzati sull'output o ancora dovuti ai processi decisionali del sistema;

step 5 assegnazione finale di flussi e capacità; i flussi finali vengono calcolati mediante algoritmi di instradamento in grado di simulare il comportamento della tecnica a coda più breve [5]; elementi correttivi sulle strategie di assegnamento dei flussi derivano da tecniche decisionali del sistema esperto.

Il passo 5 viene rieseguito per l'analisi dei casi di back-up, laddove, cioè, va esaminato il comportamen-

Fig. 1 - a) Disegno topologico iniziale; b) Percorso dei portanti fisici; c) Disegno topologico finale.



to della rete a seguito della disabilitazione di una dorsale tra 2 nodi.

L'elaborazione completa fornisce la topologia di rete, gli instradamenti di traffico, la capacità di linea, il ritardo globale medio di rete e quello lungo i diversi cammini tra periferia e host.

Il calcolo dei ritardi viene effettuato utilizzando il modello a code di Kleinrock per reti a commutazione di pacchetto [1].

4.2 Strategie decisionali

Le strategie decisionali del sistema esperto hanno un ruolo fondamentale soprattutto nella definizione della topologia e degli instradamenti.

Di seguito vengono esposte alcune problematiche risolte da tali tecniche decisionali così come le esprime l'esperto:

a) Fattore k di connettività.

Il fattore k è un parametro che entra in gioco quando si debba determinare la topologia iniziale della rete e rappresenta il numero di connessioni che un nodo deve avere con la rete stessa. La scelta del valore di k per ciascun nodo dipende dai seguenti elementi in ordine di importanza:

1) valutazione a priori dell'esperto; nella fase di allocazione dei concentratori l'esperto ha la possibilità di assegnare delle priorità a località candidate quali sedi di nodo. Tali priorità contribuiscono, nella fase successiva, alla determinazione di k;

2) nodo sede di host; con questa terminologia si indica il Front End Processor dell'host, per il quale il livello di connettività è assai elevato;

3) numero delle linee di rete secondaria; il fattore k è proporzionale al numero di linee della rete secondaria: tale condizione presuppone infatti una certa quantità di traffico che aumenta al crescere degli apparati terminali connessi e che va quindi smaltita su più cammini alternativi;

4) numero degli host cui è diretto il traffico entrante al nodo; nel disegno topologico di una rete, la tendenza è quella di generare connessioni dirette tra nodo e host, soprattutto laddove il traffico è abbastanza elevato, allo scopo di contenere i tempi di trasmissione end-to-end.

5) distanza del nodo dagli host; la distanza tra nodi e host fornisce una stima dell'ampiezza della rete: minore è tale ampiezza, minore risulta la possibilità di determinare cammini alternativi e quindi più basso è il valore di k.

b) Topologia e affidabilità.

Una volta determinata la soluzione proveniente dall'algoritmo Cut Saturation, a seconda delle dimensioni della rete in termini di ampiezza geografica, traffico, tipo nodi (di piccola, media, grande capacità), va verificato da parte del sistema esperto l'adeguamento dei cammini alternativi relativamente a portanti fisici distinti.

Ad esempio (fig. 1a-1b), il triangolo Milano-Bologna-Roma su portanti analogici non garantisce un reale stato di affidabilità: infatti le dorsali Milano-Bologna e Milano-Roma seguono lo stesso cammino (lungo l'autostrada del sole) fino alla centrale telefonica di Bologna. Se in tale sede o lungo il tratto Milano-Bologna si verificasse un guasto, Bologna stessa, ma anche Roma, rimarrebbero completamente disconnesse. Si supponga (fig. 1c) che in rete siano previsti dei nodi anche a Genova e Verona: in tal caso i 2 triangoli Milano-Bologna-Verona e Milano-Genova-Roma fornirebbero sicuramente maggiori garanzie di affidabilità.

Proprio sulla base di queste considerazioni, il sistema esperto,inglobando nella sua banca dati informazioni sullo schema di connessioni delle linee analogiche e digitali fornite da SIP e Azienda di Stato (ASST), è in grado di aggiornare la topologia onde non incorrere negli inconvenienti su esposti.

c) Aggiornamento dei flussi.

Quanto alla distribuzione dei traffici sulle diverse linee, va osservato che gli algoritmi normalmente utilizzati, quali il Flow Deviation o l'Extremal Flow, non sempre forniscono un comportamento reale del nodo: ad esempio (fig. 2), se il traffico dal nodo A al nodo B sui diversi cammini di distribuisce con certe percentuali, non è possibile che il traffico inverso (da B ad A) si riversi completamente su un solo cammino, come risulta invece dall'applicazione dei suddetti algoritmi.

Tali risultati analitici provengono da metriche di cammini-minimi: queste conducono ad ottimi teorici che poco si adattano alle reali strategie di instradamento, adottate ad esempio dai sistemi Datanet della Honeywell Bull.

Il sistema esperto deve, quindi, localizzare, alla fine di tutto il processo algoritmico, situazioni del tipo sopra esposto (nel caso di cammini multipli tra 2 end point e traffico bidirezionale tra gli stessi) e correggerle opportunamente mediante i criteri:

1) se su un certo cammino il traffico in una direzione risulta molto basso (al di sotto di una data tolleranza) e nella direzione opposta è nullo, l'instradamento su quel cammino viene eliminato;

2) se su un certo cammino il traffico in una direzione risulta elevato (superiore ad una data tolleranza) ed è nullo nella direzione opposta, in tale senso vengono ricalcolate le percentuali di instradamento, sulla base di un ritardo sui vari cammini circa uguale per le diverse quantità di traffico (fig. 3a-3b).

Conclusioni

Si è voluto qui illustrare un'applicazione dei Sistemi Esperti così come è "vista" dall'esperto che fornisce le specifiche, senza affrontare, quindi, le problematiche di costruzione della base della conoscenza e le implicazioni implementative.

Il progetto è estremamente ambizioso anche perchè, per il momento, non supportato da esempi analoghi, almeno in Italia, data soprattutto la vastità dell'area che vuole coprire e la mancanza di modelli e algoritmi esaustivi.

Il Sistema Esperto WANESSA, attualmente in fase d'implementazione, è previsto completamente funzionante nel 1989, dopo una serie di test, comprensivi anche di strumenti di simulazione, che consentiranno un collaudo approfondito e l'aggiustamento delle strategie decisionali.

Fig. 2 - Distribuzione errata del traffico fornita da algoritmi correnti.

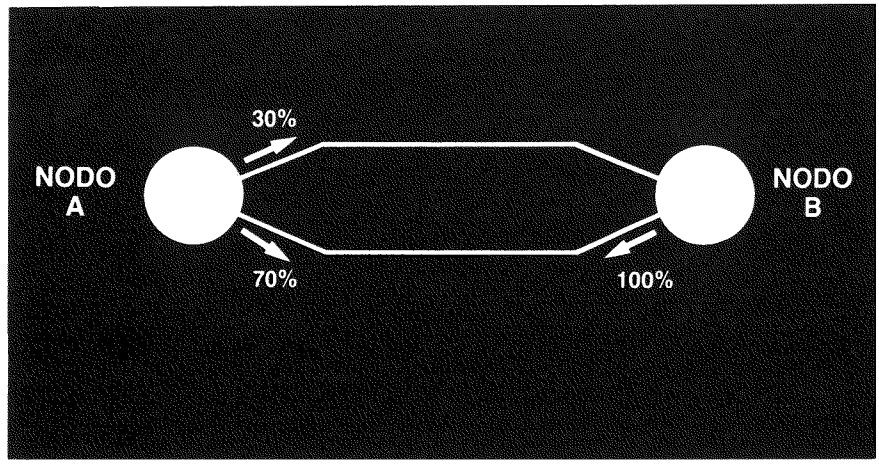
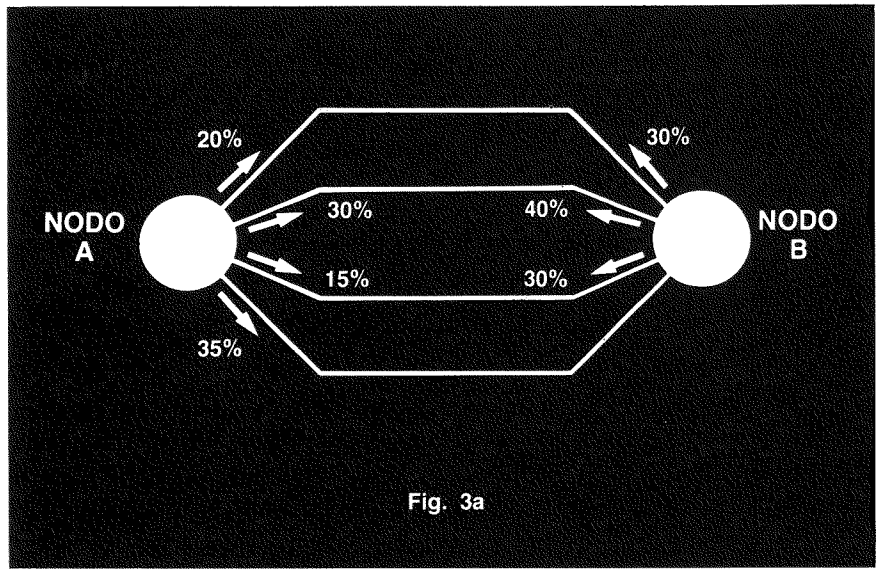
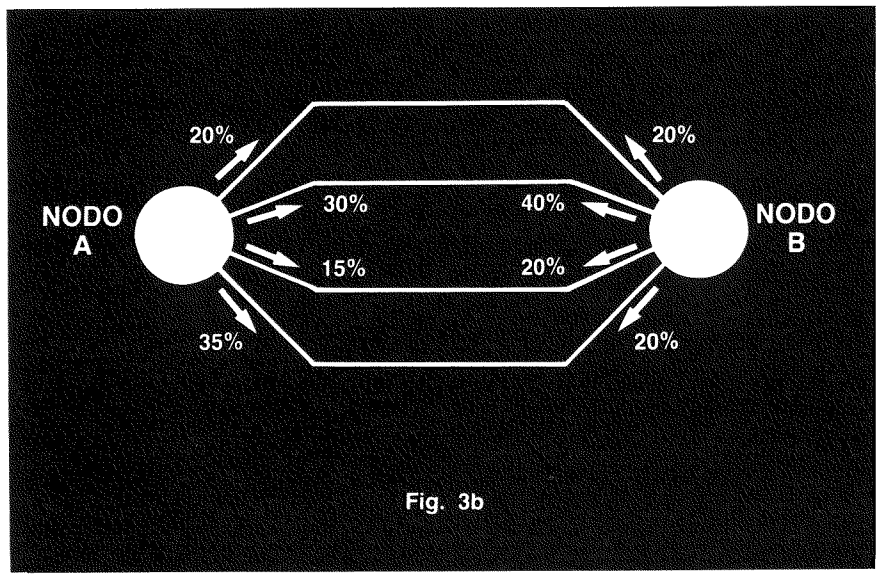


Fig. 3 - a) Distribuzione del traffico fornita da algoritmo; b) Ridistribuzione del traffico effettuata dal Sistema Esperto.





Referenze

[1] - R.R. BOORSTYN and H. FRANCK, *"Large-Scale Network Topological Optimization,"* IEEE Transactions on Communications, vol. COM-25, pag. 29-47, Gen. 1977.

[2] - A.S. TANENBAUM, *"Computer Networks"*, Prentice Hall, 1981.

[3] - M. GERLA, L. FRATTA and L. KLEINROCK, *"The Flow Deviation Method: An Approach to Store-and-Forward Communication Network Design"*, Networks, vol. 3, pag. 97-133.

[4] - M. GERLA, H. FRANK, W. CHOU and J. ECKL, *"A Cut Saturation Algorithm for Topological Design of Packet Switched Communication Networks"*, Proceeding of National Telecommunication Conference, S. Diego 1974, pag. 1074-1085.

[5] - R.G. GALLAGER, *"A Minimum Delay Routing Algorithm Using Distributed Computation"*, IEEE Transactions on Communication, vol. COM-25, pag. 73-85.

Interazione vocale col computer

JOHN WATERWORTH

British Telecom Research Labs.
Ipswich (Gran Bretagna)

1. Introduzione

La parola sembra presentare un certo numero di vantaggi come mezzo di comunicazione; è naturale e quindi è familiare, si può desumere essere conveniente ed il suo uso con i computer potrebbe, in teoria, convertire ogni telefono in un terminale remoto a basso costo. La gente potrebbe quindi accedere a tutte le informazioni e servizi oggi resi disponibili dai sistemi computerizzati per mezzo di un umile ma onnipresente telefono. La parola ha anche indiscussi vantaggi in certe situazioni dove altri strumenti sono impossibili o non convenienti ad usarsi per particolari categorie di utenti. Può essere impiegata in ambienti ostili, dove la visibilità è limitata e in circostanze che rendono desiderabile o necessario avere le mani e gli occhi liberi per altri compiti (ad es., sistemi di allarme). Analogamente, anche persone con handicap visivi o fisici possono trarre vantaggi dal mezzo così come gli analfabeti (attualmente circa 5 milioni nel Regno Unito). Altre applicazioni, anche se banali in confronto, sono per individui che preferiscono non usare altri mezzi di interazione disponibili e, più semplicemente, anche solo per il gusto della novità (telefoni giocattolo per manager, macchine per il caffè parlanti, ecc.).

Qualunque sia il motivo della scelta della parola come interfaccia, il suo uso condurrà inevitabilmente a mutamenti nelle relazioni tra utenti e sistemi. Parlare ed ascoltare sono attività che sono intimamente connesse con le convenzioni sociali della comunicazione interpersonale acquisita durante il nostro sviluppo. Premere bottoni e guardare schermi innescano un diverso insieme di associazioni e comportamenti.

A causa di ciò, la comunicazione vocale uomo/macchina genera diversi problemi che rendono questa un'area particolarmente fertile per ricerche sui fattori umani. Dal momento che la parola è un canale così naturale, l'utente porterà alle interazioni vocali una quantità di esperienze ed aspettative non connesse alla comunicazione non vocale. La parola è naturalmente usata per moltissimi scopi (ad es., per evitare attività peggiori, per conformarsi alle norme, per esprimere affetto, ecc.), oltre che per lo scambio di informazioni fattuali (Robinson, 1972). Cioè, la maggior parte del parlato naturale esegue funzioni che sono importanti nella comunicazione tra le persone (per es., farsi delle amicizie, esprimere sentimenti ed emozioni, raccontare barzellette, sgridare), ma che potrebbero non essere adatte per l'interazione uomo/macchina (almeno nell'immediato futuro).

Un altro fattore cruciale è la natura temporanea del parlato ed il modo con cui ciò interagisce con le limitazioni della capacità umana di elaborare informazioni. La parola è efemerica, non può essere ricercata e così deve essere tenuta in memoria per un tempo sufficientemente lungo per essere interpretata ed usata. Ma la quantità di informazioni che gli umani possono attivamente conservare in tale modo è molto limitata e questa ridotta capacità deve spesso essere condivisa dalle operazioni cognitive in corso. Può essere che tenendo in considerazione queste due caratteristiche della parola, si siano sviluppate molte convenzioni per controllarne l'uso nelle interazioni naturali (ad es., le regole di apertura, chiusura e scambio: ved. Schegloff e Sacks, 1973), che permettono al parlatore/ascoltatore d'ottenere un interscambio fluente, dando il modo di controllare lo stesso processo conversazionale (per es., la regolazione di incontro). Può essere necessario tener conto di queste convenzioni quando si progettano sistemi interattivi basati sulla parola. Ciò comporterà il dover dotare il sistema di un modello di prestazione conversazionale, come pure di mezzi per consolidare l'attuale stato dell'interazione.

2. La parola come mezzo per comunicare

La capacità di capire e produrre un linguaggio parlato è acquisita naturalmente e senza sforzi dalla maggior parte delle persone durante i primi 3 anni di vita e, con la pratica e l'esercizio, diventa una specializzazione che fornisce un efficiente

(*) Il materiale di questo articolo è tratto dal volume: A. Monk (a cura di), *Interazione uomo-computer,* della collana dei Quaderni di Informatica (1988).

mezzo per comunicare. Si dovrebbe notare che, anche se molti imparano a leggere, ciò non avviene naturalmente.

Sembra che la parola sia un codice evolutosi per sfruttare con grande efficacia le nostre basilari capacità anatomiche, fisiologiche e cognitive ed è quindi un ottimo candidato da studiare, se desideriamo scoprire i mezzi più naturali ed efficienti per gli umani per comunicare l'un l'altro, o con le macchine.

Capire e produrre parole è normalmente così facile, che non ci rendiamo conto dei processi che accompagnano la conversione di un'idea nella sua rappresentazione linguistica in un'onda acustica, o la ricostruzione dell'idea espressa in un'onda acustica che arriva alle nostre orecchie. Eppure evidentemente la rappresentazione dell'informazione in un messaggio linguistico parlato assume forme diverse. La possiamo considerare in termini di comandi neurali ai muscoli del tratto vocale, come movimenti del tratto vocale e delle sue articolazioni (lingua, labbra, mascelle, ecc.), come i suoni vocali che ne risultano, come gli schemi di attività dei nervi che portano segnali dell'orecchio al cervello, o in termini più astratti, come una rappresentazione psicologica interna di parole, dei loro significati e delle loro relazioni sintattiche in una frase.

Questa molteplicità di livelli di rappresentazione ha condotto a considerare il processo di comprensione vocale come un insieme di sottoprocessi, ciascuno dei quali trasforma l'informazione in un messaggio da una rappresentazione ad un'altra. Uno schema di questo tipo è illustrato in fig. 1.

Descrivere la produzione e la comprensione di parole (o ogni altro processo psicologico) come un insieme di fasi discrete di elaborazione di informazioni è conveniente perchè implica che ciascuna fase può essere studiata in modo relativamente indipendente dalle altre, ma è importante sapere che questo è solo un modello descrittivo semplificato del processo vocale. Il modello dovrà essere modificato alla luce dei risultati di esperimenti progettati per identificare e caratterizzare ciascuna fase e non abbiamo a priori alcun motivo di supporre, ad esempio, che l'elaborazione delle informazioni nel cervello proceda nel modo strettamente gerarchico indicato dal modello di fig. 1. Ciò nonostante, i processi psicologici di creazione di modelli, spesso con modelli espressi sotto forma di programmi per computer, possono essere una ricca sorgente di idee per esperimenti e teorie, purché si ricordi che i modelli psicologici possono essere validi solamente con dati ottenuti da esperimenti psicologici.

3. Parlato generato da macchine

Gli annunci di un orologio parlante fanno uso del principio di concatenazione (ved. cap. 12, par. 3), per selezionare in modo meccanico il necessario vocabolario in modo da dare il messaggio appropriato ogni 10 secondi. Questo principio viene esteso, negli scambi del sistema X, alla concatenazione in tempo reale di annunci codificati in modo digitale, opportuni per l'uso di "star services" (quali chiamata semplificata, servizi di deviazione, ripetizione dell'ultima chiamata, ecc.).

Questo metodo ha il vantaggio di dover ri-registrare solo una piccola parte del vocabolario per aggiornare gli annunci, ciascun utente sente ogni annuncio dall'inizio ed i messaggi possono essere personalizzati per servire ai bisogni di singoli utenti. A causa di questi vantaggi, un tale sistema può anche essere usato per i servizi interattivi di informazioni generali.

3.1 Atteggiamenti e aspettative

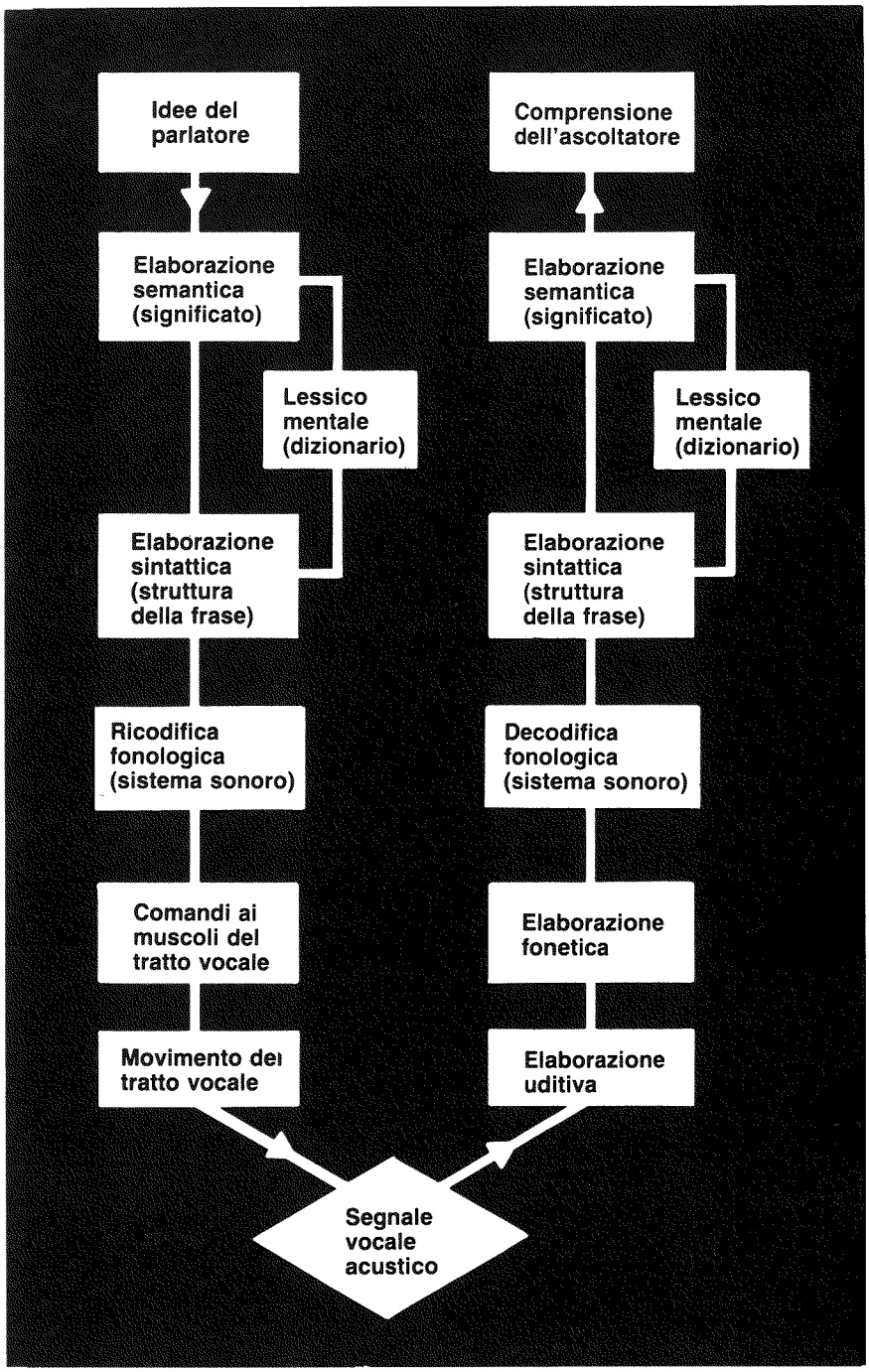
La voce dell'annunciatore usata per generare gli annunci ha un effetto sugli atteggiamenti e sulle aspettative dell'utente nei riguardi del sistema. La scelta dell'annunciatore per la registrazione originale dell'annuncio è stato il risultato di indagini di preferenza del pubblico effettuate su un vasto campione di volontari dalla British Telecom a Ipswich. Si è anche fatto un tentativo di correlare la preferenza risultata e la personalità ascritta (Cox e Cooper, 1981). In un esperimento, i soggetti hanno votato sulla preferenza di diverse voci e su 28 scale di personalità bipolare (ad esempio, serena - irritabile, intelligente - stupida). L'analisi di fattore ha diviso le votazione di personalità in 2 principali fattori - piacevolezza e arroganza. In generale, gli annunciatori femmina hanno ottenuto un leggero vantaggio sui maschi nelle votazioni di preferenza e sono state segnalate positivamente per una manifesta arroganza, mentre gli uomini sono stati segnalati negativamente per questa stessa ragione. Questi risultati, naturalmente, forniscono solo ragioni molto parziali del perché una voce di annunciatore può essere preferita ad un'altra. Il contenuto degli annunci è fondamentale e si è posto un considerevole impegno per mettere a punto forme chiare e concise di parole.

La qualità percepita di parlato concatenato di forme d'onda memorizzate può essere molto alta. È possibile migliorarla anche di più, sulla base di uno studio dettagliato della prosodia del linguaggio naturale, ma il costo di realizzazione sarebbe alto se, come probabilmente avverrebbe, si dovessero memorizzare versioni plurime di ciascun componente di concatenazione. L'introduzione di una struttura ritmica più naturale potrebbe tuttavia ottenersi a costi minimi. Ma non dovremmo sottovalutare il pericolo che una pronuncia molto naturale potrebbe essere scambiata per un annunciatore vivente, con tutta la confusione che ciò potrebbe causare. Esperimenti recenti hanno mostrato che le caratteristiche della voce scelta per un sistema influenza il modo con cui gli utenti tendono ad interagire. Una voce dal timbro meccanico sembra incoraggiare l'attribuzione di un più basso livello di intelligenza che non una pronuncia che suoni più naturale.

3.2 Obblighi di elaborazione dell'utente

L'uso del canale uditivo per la presentazione delle informazioni all'utente fa sorgere diversi problemi per il progettista di dialogo. Dal momento che il parlato non può es-

Fig. 1 - Modello descrittivo del processo di comunicazione vocale.



sere ricercato, l'informazione deve essere conservata in una memoria di lavoro per un tempo sufficientemente lungo per la sua interpretazione ed il suo uso. Baddeley e Hitch (1974) hanno ipotizzato che la memoria di lavoro comprenda due componenti: un anello a capacità molto limitata e un elaboratore centrale che ha anche qualche capacità di memorizzazione. La dimensione della memoria nell'anello si pensa sia limitata a circa 3 campi e può essere utilizzata senza porre richieste sostanziali all'elaboratore. Beddeley e Hitch hanno scoperto che i compiti di elaborazione quali il ragionamento verbale, la comprensione di prosa e l'apprendimento a ricordo libero erano impediti quando venivano trattenuti nella memoria di lavoro più di 3 campi, ma non quando ve ne erano 3 o meno di 3, perché nell'ultimo caso veniva coinvolto solo l'anello. Una ricerca successiva ha dimostrato che l'anello sembra avere limiti temporali, piuttosto che sul numero di campi, essendo in grado di contenere solo tanto materiale quanto può essere articolato in circa 2 secondi. Invece, l'elaboratore centrale è probabilmente basato sui campi piuttosto che sui tempi (Baddeley, 1976). Il cap. 3 fornisce una descrizione delle capacità ipotetiche della memoria umana e delle loro caratteristiche.

Abbiamo esaminato l'impatto di questi limiti umani sull'assimilazione di informazioni vocali usando, per esempio, menu di dimensione e di complessità concettuale variabili.

In una ricerca sono stati messi a confronto due sistemi informativi usanti menu uditivi. Un sistema semplice, che forniva informazioni su un problema molto limitato, non ha creato alcuna difficoltà agli utenti. Qui la stuttura del menu era molto chiara; non vi erano categorie sovrapposte o strutture ad albero per introdurre un elemento risolutorio nel problema.

Una versione uditiva di Prestel, d'altra parte, si è dimostrata molto difficile per gli utenti. Essi spesso hanno richiesto la ripetizione del menu e sono stati spesso incapaci di comprendere la struttura dei menu mentre mantenevano le opzioni in memoria. Cioè, le necessità di risoluzione e di memoria combinate sono parse superare la capacità delle memorie di lavoro degli utenti. Gli stessi menu di Prestel in modalità visiva non avrebbero probabilmente causato problemi. Il menu visivo può essere consultato ripetutamente, cosicché il carico di memoria è ridotto e segnali visivi come il colore possono fornire indicatori verso la struttura di menu, alleviando quindi in un certo modo la necessità risolutoria.

La presentazione di stringhe di caratteri, tramite parlato concatenato, ha generato problemi particolari Il confronto della memorizzazione a breve termine di stringhe di caratteri annunciate automaticamente con quelle pronunciate da un operatore umano hanno indicato prestazioni inferiori nel primo caso.

È stata condotta una ricerca (Waterworth, 1983) per esplorare gli effetti del tempo e dell'intonazione sulle preferenze degli utenti e sulla prestazione di memoria con stringhe di caratteri.

Si è trovato che la prestazione di memoria potrebbe migliorare significativamente aumentando la durata delle pause tra sottogruppi di caratteri nell'ambito di una stringa, confermando il lavoro di Grusec (1976), anche se i giudizi di preferenza non ne sono stati influenzati. Di converso, usando uno schema di intonazione più sofisticato e naturale, si è migliorato in modo significativo la preferenza soggettiva, ma non è stata influenzata la capacità di memorizzazione. Non vi sono neppure stati effetti di interazione tra la durata e l'intonazione, anche se è noto che questi fattori interagiscono nella produzione di parlato naturale (Pike, 1945).

Un'alternativa alla concatenazione di elementi vocali a forma d'onda memorizzata controllata dalla macchina è di sintetizzare il parlato a partire da un insieme di parametri. Questo metodo richiede una memoria del sistema molto inferiore delle tecniche a forma d'onda memorizzata ed elimina la necessità di registrare una voce umana. Rende disponibile un vocabolario virtualmente illimitato e può conseguentemente fornire la piacevole possibilità di tradurre un testo direttamente in un parlato (cap. 12, par. 3). Queste caratteristiche danno un enorme flessibilità al parlato automatico che può essere usato a sostegno di un'ampia gamma di servizi, a partire da insiemi limitati di annunci, come nel caso di Star Services, a servizi informativi interattivi completi, con banche dati ampie e variabili.

Sfortunatamente, gli algoritmi attualmente disponibili per la generazione vocale da input codificato producono risultati sonori abbastanza innaturali.
L'uso della parola sintetizzata esaspera i problemi relativi alla presentazione dell'informazione mediante la voce. Luce *et al.* (1983) hanno dimostrato che la parola sintetizzata richiede un impegno elaborativo più grande sulla memoria di lavoro che non il parlato naturale e che ciò è in parte responsabile delle difficoltà nella sua percezione e comprensione.
Si ipotizza che ciò possa essere dovuto alla relativa mancanza di informazioni ridondanti nella parola sintetizzata. Non è ancora completamente chiaro, tuttavia, se i ridotti livelli di prestazione con le parole sintetizzate non siano dovuti a pura mancanza di dimestichezza con esse.
Aumentare la ridondanza del materiale può compensare la minore intelligibilità, fino a un cerpo punto (anche se ciò ovviamente rallenterà la velocità dello scambio di informazioni), e l'intonazione realistica può anche fornire segnali all'ascoltatore. Un esempio di un messaggio ad alto grado di ridondanza potrebbe essere qualcosa come *"È necessario che tu identifichi il tuo conto con l'introduzione del numero di conto. Per favore introduci tutti gli 8 caratteri del tuo numero di conto bancario premendo i tasti opportuni sul tuo telefono ora"*. Un messaggio simile con poca ridondanza sarebbe "Introduci ora il numero di conto".
Un recente esperimento ha dimostrato che la scelta del vocabolario usato nei messaggi può avere un notevole effetto sull'intelligibilità, ma che non è possibile compensare i problemi di parola semplicemente rallentando la velocità di pronuncia (ved. cap. 7). Nel parlato naturale la velocità varia con la ridondanza del materiale, cosicché messaggi ad alta ridondanza vengono normalmente pronunciati più rapidamente di quelli con una ridondanza relativamente bassa. Ciò può essere un tentativo di ottimizzare lo scambio di informazioni mantenendo la velocità del processo costante per entrambi, parlatore e ascoltatore. Un sistema sofisticato potrebbe essere in grado di regolare la velocità di pronuncia in questo modo. Ciò certamente migliorerebbe la naturalezza anche se non è certo che l'accuratezza della ricezione dell'informazione ne verrebbe migliorata. La velocità non sarà un problema per gruppi di annuncio relativamente semplici, anche se si dovrebbe ricordare che la velocità influenza caratteristiche di personalità attribuita quali competenza e benevolenza (Smith *et al.*, 1975). Le caratteristiche prosodiche sono molto significative per parlato naturale non vincolato e possono essere usate per manifestare intenzione, emozione, interpretazione, elaborazione ed anche negazione del significato letterale. La durata delle pause, ad esempio, ha significati semantici e questo suggerisce che i tipi di ritardo che si possono sperimentare con interazioni al videoterminale/telescrivente non possono essere tollerate nel parlato. Se possibile, le pause dovrebbero essere limitate alle sospensioni naturali nello scorrimento verbale. È possibile che un inappropriato stile

di pronuncia possa generare attribuzioni di personalità o significati nascosti che potrebbero ostacolare il processo di scambio di informazioni.

4. Riconoscimento vocale

L'uso di apparecchiature di riconoscimento vocale (riconoscitori automatici vocali o ASR) offre diversi vantaggi rispetto all'input di tastiera, principalmente perché le parole comando possono essere scelte per combinarsi con le funzioni a cui danno l'avvio, cosicché non vi è alcuna necessità per l'utente di schematizzare le funzioni mediante simboli arbitrari, ma la tecnologia disponibile non consente attualmente interazioni particolarmente confortevoli tra uomo e macchina, nonostante la dichiarata naturalezza e convenienza del mezzo.

Il progetto del dialogo deve tener conto del fatto che la parola "riconosciuta" sarà spesso sbagliata, cosicché l'utente avrà spesso bisogno di confermare o correggere l'input che la macchina ha identificato.

4.1 Accuratezza nella misura del riconoscimento.

La scelta del riconoscitore più opportuno per un particolare compito può presentare diverse difficoltà. I produttori citano normalmente valori di accuratezza fra il 95 ed il 99% per i loro prodotti, di qualunque tipo essi siano. Ma ciò ha ben poco significato se non si considerano altre informazioni. Tali valori sono presumibilmente ottenuti mediante una scelta molto accurata su un piccolo dizionario facilmente distinguibile e mediante prove in condizioni ideali con utenti esperti. L'accuratezza di riconoscimento (in termini di numero di successi sul numero di tentativi) è importante, ma sono critiche le circostanze in cui i risultati sono ottenuti e l'accuratezza non è tutto; anche la confusione deve essere considerata. Si è scoperto che le matrici degli alberi di confusione sono utili nel selezionare vocabolari funzionali per particolari applicazioni. Un albero di confusione dei numeri da 1 a 5 e di 5 parole comando, ottenuto da un poco costoso riconoscitore di parole singole, è illustrato nella fig. 2. I risultati provengono da 60 soggetti, ciascuno che ripete tutte le 10 parole 20 volte (cioè 1200 pronunce di ciascuna parola). Certe parole sono confuse molto frequentemente (per esempio, 1 e 5, 3 e *repeat*), mentre per altre questo è raramente il caso (per esempio 1 e *repeat*).

Ciò che l'albero di confusione non mostra, mentre lo mostrerebbe una matrice di confusione, è che lo schema è spesso asimmetrico. In qualche tentativo, *repeat* è stato compreso come 3 oltre tre volte più frequentemente di quanto 3 sia stato riconosciuto come *repeat.*

La fig. 3 presenta un albero di confusione per i numeri da 1 a 10 ottenuto da 24 soggetti, ciascuno che ripete tutte le 10 parole 32 volte (768 pronunce di ciascuna parola). Ciò indica chiaramente che certi numeri, nella fattispecie 5 e 9, non dovrebbero essere usati assieme con questa apparecchiatura, se fosse possibile. Se si è dimostrato necessario riconoscere tutti i numeri, allora gli utenti dovebbero essere incoraggiati a pronunciare male le parole per renderle più distinguibili. Riflessioni sul probabile schema di confusio-

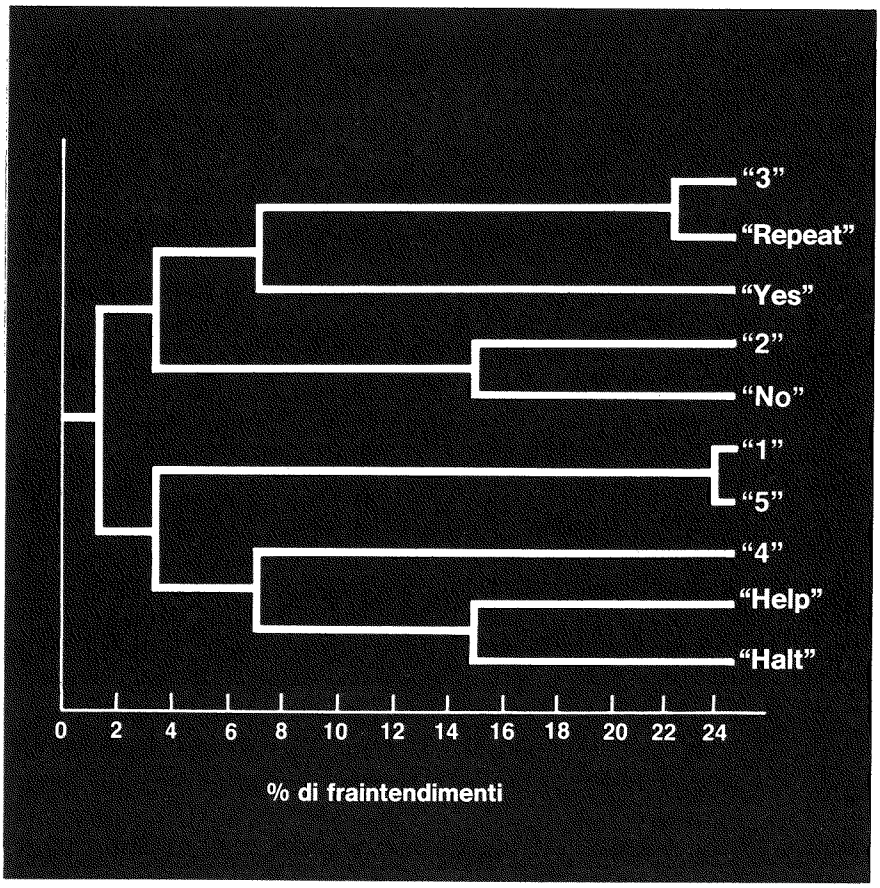


Fig. 2 - Albero di confusione per i numeri da 1 a 5 e cinque parole comando.
I campi collegati sulla destra dell'albero vengono facilmente confusi (per esempio 1 e 5), i campi collegati sulla sinistra non sono facilmente confusi (per esempio NO e 1).

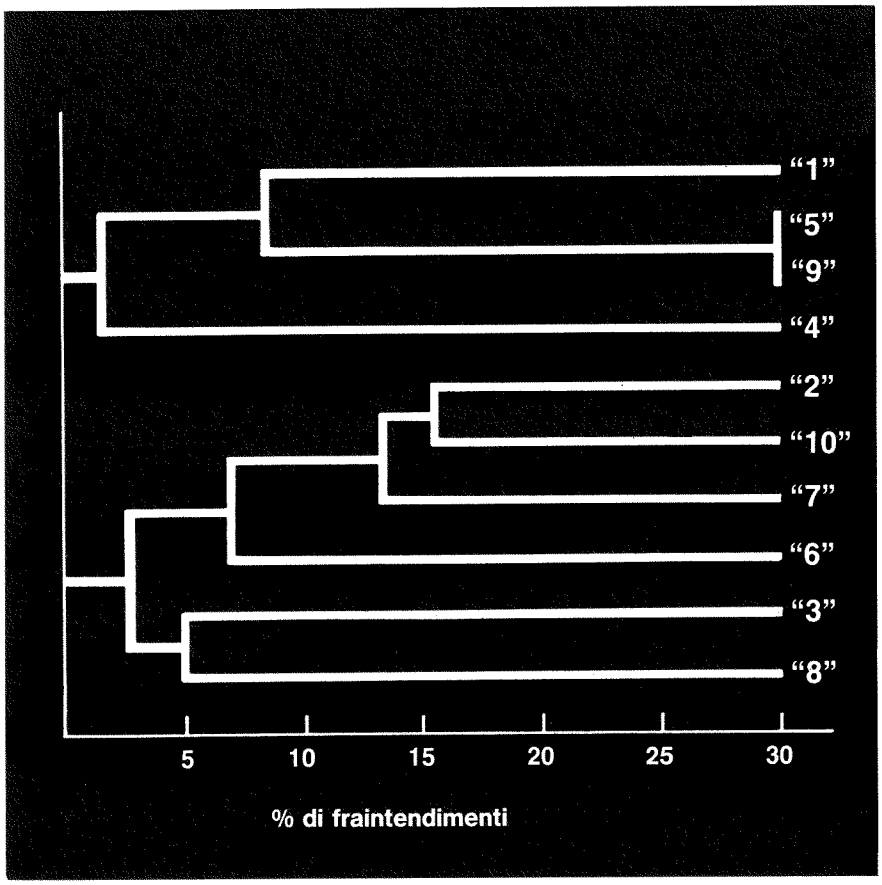


Fig. 3 - Albero di confusione per i numeri da 1 a 10.

ni del dizionario sono essenziali per effettuare scelte corrette di parole comando per una data applicazione.

I riconoscimenti devono essere valutati sul vocabolario che si intende utilizzare, magari usando una procedura analoga a quella descritta per la messa a punto dei sintetizzatori vocali (ved. cap. 7). Anche preferibile è provare la macchina nell'applicazione da svolgere, ma questo non può essere fatto prima di acquisire l'apparecchiatura. Un approccio che sta diventando comune è di fornire nastri di vocabolario standard per una gamma di diversi tipi di macchine. I produttori dovranno inevitabilmente mettere a punto i loro sistemi per farli operare con successo sui nastri di prova, in modo tale che la prestazione dichiarata non possa essere ottenuta nell'uso generale senza una scelta accurata del materiale registrato sui nastri.

4.2 Limitazioni imposte dai sistemi esistenti.

I sistemi di riconoscimento attualmente disponibili impongono un gran numero di limitazioni alle forme che l'interazione uomo-macchina può assumere. Per es., sono normalmente solo capaci di riconoscere parole pronunciate separatamente e spesso solo da un parlatore che abbia preventivamente addestrato il sistema su un dizionario particolare e molto limitato. Per il riconoscimento di parole isolate, il sistema deve interrogare l'utente in modo tale da sollecitare una pronuncia di risposta ammessa. Ciò significa che occorre porre un'estrema attenzione al progetto di dialogo vocale uomo-macchina, per ottenere dall'utente risposte appropriate.

Molte apparecchiature esigono che l'utente ripeta il vocabolario di parole da riconoscere parecchie volte, per addestrare il riconoscitore. Il tempo e l'impegno richiesti per addestrare un riconoscitore ad una particolare voce sono probabilmente così grandi da impedire l'uso del sistema. È stato ipotizzato che il "mascherare" la sessione di addestramento solleciti pronunce più simili a quelle prodotte durante l'uso reale che non quando si effettuano sessioni di addestramento formali, anche se non è stato verificato il successo di questo approccio.

Il metodo dipende dalla prevedibilità delle pronunce dell'utente, mentre quest'ultimo ne ricava l'impressione che egli stia attualmente decidendo cosa dire. Per l'utente la distinzione tra addestramento e uso deve scomparire. Se ciò può evitare una parte delle irritazioni provocate dalle sessioni di addestramento, è più dubbio, dal punto di vista dell'utente, quanto l'addestramento nascosto assomigli veramente alla situazione reale. È molto difficile immaginare compiti di addestramento che imitino realisticamente l'applicazione effettiva e diano l'impressione di controllo soggettivo, mentre sono, in effetti, completamente determinati. Questi compiti di addestramento tendono ad essere banalmente facili per gli utenti, a differenza dell'utilizzo effettivo del sistema, per cui il dichiarato vantaggio di questa tecnica, che cioè pronunce simili sono prodotte in fase di addestramento e nell'uso effettivo, può essere più apparente che reale.

Anche la sollecitazione vocale ha delle implicazioni in quest'area. Abbiamo osservato che gli utenti che stanno addestrando un riconoscitore a rispondere a sollecitazioni vocali prodotte sinteticamente hanno la tendenza ad imitare il modo con cui tali sollecitazioni sono pronunciate. Se questo è vero anche durante l'uso effettivo del sistema, il riconoscimento potrebbe essere migliorato. Ma se l'utente non imita la voce di sollecito anche durante l'uso del servizio, la qualità del riconoscimento ne risulterà degradata. Gli effetti delle sollecitazioni visive nei confronti di quelle uditive necessitano di prove. La pratica di una sessione molto breve di addestramento, per sintonizzare un particolare utente su forme pre-programmate, aumenterà probabilmente l'accettabilità futura di dispositivi dipendenti dall'utente.

Per un esiguo gruppo di utenti regolari e motivati, la necessità di addestrare il sistema e di aggiornare occasionalmente le forme, non è un grosso problema. Naturalmente, l'obiettivo finale deve essere lo sviluppo di tecniche che non richiedano alcun addestramento del sistema da parte dell'utente e questo è particolarmente importante per i servizi via telefono usati dal pubblico. Non si può richiedere all'utente di addestrare il sistema sulla sua propria voce ogni volta che desidera conoscere l'orario di un autobus o le condizioni metereologiche. I riconoscitori che sono disponibili oggi, dichiarati indipendenti dal parlatore, sono stati in effetti addestrati da un ampio gruppo di utenti a produrre diverse versioni di ciascuna forma di parola, o a fondere molte pronunce in forme composite. Sfortunatamente, nessuno dei due approcci è molto soddisfacente e il miglior modo di produrre forme di uso generale non è noto. La British Telecom sta attualmente esplorando un approccio a questo problema, quello di usare diversi esemplari di tipi di voce come forme, da cui potrebbe essere selezionato per ciascun utente un particolare insieme sulla base della sua prima pronuncia. Tuttavia, non è probabile che qualsiasi sistema sarà in grado di districarsi con l'intera gamma di accenti e dialetti inglesi, almeno nel futuro prevedibile. E vi sarà sempre uno strato di popolazione per la quale il riconoscimento vocale automatico non è possibile, da cui la necessità di intervento di un operatore in casi di seria difficoltà.

È ora disponibile il riconoscimento di parole collegate, anche se ciò ancora comporta normalmente addestramento su singole parole discrete. I problemi connessi con l'uso di un riconoscitore possono effettivamente aumentare con il riconoscimento di parole collegate, perché il parlato collegato naturale non consiste di una sequenza di parole intere complete e non sovrapponentesi. Naturalmente, l'utente riuscirà soltanto se usa parole su cui il sistema è stato addestrato e dovrà parlare in un modo molto artificiale.

È difficile trovare un modo semplice per incoraggiare l'utente a fornire risposte accettabili. È anche molto più difficile dare all'utente un modello utile di come il sistema opera di quanto non avvenga nel caso di macchine a parola singola.

Il concetto di un riconoscitore vocale che riconosce tutto di qualche frase, parti di altre e nulla di qualcuna non è affatto ovvio.

Un altro maggior svantaggio è che questo approccio non è estendibile praticamente al riconoscimento di parlato continuo da un discorso senza vincoli, a causa degli obblighi lessicali.

5 Sistemi interattivi

5.1 Il processo di progettazione

I fattori che riguardano la generazione ed il riconoscimento vocale non dovrebbero essere visti separatamente dal processo a due vie di scambio di informazioni tra l'utente ed il sistema.

Per esempio, le voci sintetiche che sembrano inaccettabili in esperimenti di ascolto molto rigidi possono essere meglio accolte nel contesto dell'uso di un servizio reale. Il riconoscimento vocale, d'altra parte, ha normalmente minor successo in situazioni libere che non in ambienti controllati con un parlatore che pronuncia in modo accurato.

Il nostro lavoro in quest'area include l'avanzare raccomandazioni specifiche per il progetto di particolari sistemi interattivi, l'usare input sia da tastiera che da voce e il valutare questi sistemi in condizioni d'impiego con rappresentanti del nostro gruppo di soggetti. Sulla base di una valutazione iniziale, i sistemi vengono modificati e quindi riverificati. Con questo approccio, lo sviluppo del sistema diviene un ciclo interattivo di modifiche e di valutazione del progetto, che ci consente di sviluppare una metodologia di valutazione dei sistemi e anche una "sensibilità" per i tipi di strutture di dialogo probabilmente accettabili per particolari applicazioni.

5.2 Caratteristiche di un dialogo uomo-computer

Anche se l'applicazione specifica avrà sempre una profonda influenza sulla struttura finale degli scambi possibili, si è fatto un tentativo di categorizzare le caratteristiche della comunicazione vocale uomo-macchina (Waterworth, 1982). Questo lavoro si è basato sulla distinzione, tratta da Bunt *et al.,* (1978), tra due tipi di "atti di dialogo": quelli orientati all'obiettivo e quelli di controllo dialogo. I primi implicano la comunicazione diretta tra sistema e utente di informazioni significative per l'obiettivo; i secondi sono necessari nella prevenzione e correzione di errori e nel flusso generale del dialogo stesso, ad es. nel controllo del processo di scambio di informazioni.

Waterworth (1982) ha suddiviso gli atti di questi dialoghi in specifici "sottodialoghi", sulla base delle funzioni che assolvono nelle conversazioni uomo-macchina e ha suggerito i modi più opportuni di introduzione di dati per sottodialoghi particolari. Esempi di questi diversi tipi di sottodialogo sono contenuti in "Prologo", dove il sistema fornisce all'inizio un saluto e l'identificazione del sistema, informazioni sulle

possibilità delle informazioni disponibili, nonché alcune indicazioni preliminari sul suo uso e le possibilità di ripristino/riparazione, risultanti da ripetuti errori nel confermare un input, o difficoltà nel riconoscimento positivo di risposte alle sollecitazioni (fig. 4).

Waterworth ha suggerito che i modi di introduzione di dati vocali potrebbero essere ordinati gerarchicamente dai più sofisticati e meno espliciti (in termini di quanto l'utente sa delle esigenze del sistema), posti al vertice, ai più espliciti ed elementari, alla base. E, forse più polemicamente, si è sostenuto che parallela a questa dimensione ve ne è un'altra, l'ampiezza accettabile della classe di risposta utente con quei modi, con un'ampia gamma di input possibili posti nella zona alta della gerarchia.

I sistemi devono avere la capacità di scendere, o salire, ad un opportuno livello di interazione in base al comportamento interattivo degli utenti. L'approccio del "controllo di dialogo" suggerisce che un modo per minimizzare le difficoltà utente con i sistemi basati sulla voce è di consentire all'utente il controllo del processo interattivo e in particolare la presentazione delle informazioni vocali. Se egli può rallentare, accelerare, arrestarsi e ripetere annunci e ricercare in avanti e all'indietro attraverso il dialogo, i problemi di memoria che sorgono dall'uso del parlato potrebbero apparentemente essere superati. Ma il problema rimane. Ad esempio, come saranno informati gli utenti di queste possibilità e del modo di richiamarle? Ciò, di per sé, presenta problemi di trasferimento di informazioni, a meno che venga impiegata qualche altra modalità (ad es., la guida testuale). Un approccio è quello di rendere il sistema sensibile all'utente, mediante strategie quali la misura del tempo di reazione, l'individuazione delle iterazioni, ecc. L'utente in difficoltà (e quello esperto) possono quindi essere identificati dal sistema stesso e il dialogo si può svolgere in conseguenza. Questi due approcci sono attualmente complementari. Il primo comporta la dotazione di strumenti di controllo ed il dare all'utente un utile modello del sistema, cosicché possa assumere il controllo del dialogo; il secondo è effettivamente un esempio di sistema che sviluppa un modello dell'utente, in modo che possano essere date automaticamente le sollecitazioni e indicazioni opportune.

L'interazione uomo-macchina procede così verso una maggior approssimazione del modo in cui le persone parlano tra loro e controllano il processo di dialogo, sulla base di un modello che ciascuno si è formato dell'altro e di una conoscenza comune circa le convenzioni conversazionali.

5.3 Alcune semplici regole

La parola sintetica dovrebbe essere realizzata in modo tale che l'utente non abbia una visione poco realistica delle capacità del sistema. A causa della natura temporanea della parola e delle limitazioni della memoria umana, le informazioni devono essere presentate in "pacchetti" sufficientemente piccoli per essere facilmente assimilati e adoperati dall'ascoltatore. Un parlato molto innaturale tenderà a diminuire la capacità di ritenzione ben al di là di qualsiasi decremento dovuto alla scarsa intelligibilità. La ridondanza nei messaggi vocali può aiutare ed il vocabolario dovrebbe essere scelto per soddisfare le aspettative degli utenti. Idealmente, l'output vocale dovrebbe essere controllabile dall'utente per evitare un sovraccarico di memoria. Altre strategie progettuali di output vocale comprendono l'uso di una introduzione con poco contenuto informativo. Ciò consente all'utente di sintonizzarsi con il particolare sintetizzatore usato senza perdere informazioni vitali sull'uso del sistema. Voci diverse possono anche essere usate per evidenziare livelli o punti diversi nell'ambito della struttura del dialogo. Questi suggerimenti uditivi sono analoghi all'uso del grassetto o del colore nelle visualizzazioni e possono essere usati per aiutare l'utente ad ottenere un'impressione più chiara della struttura relativa all'informazione presentata.

Nell'attuale stato dell'arte, il riconoscimento vocale dovrebbe essere trattato con cautela. Le apparecchiature hanno capacità molto limitate e la struttura del dialogo, per es. la

Fig. 4 - Atti di dialogo e sottodialoghi della conversazione uomo-macchina.

scelta delle sollecitazioni di guida dell'utente verso il suo obiettivo, richiedono una considerazione molto attenta. La valutazione di un sistema impiegato da soggetti sprovveduti e la susseguente modifica, è nella maggior parte dei casi altamente desiderabile e probabilmente necessaria. I sistemi dovrebbero essere sensibili al comportamento degli utenti e capaci di adattare conseguentemente il livello dell'interazione (per es., la specificità e la complessità dei solleciti). I sistemi futuri avranno probabilmente un modello sofisticato di stati e di comportamenti conversazionali dell'utente.

6. Conclusioni

Si stanno ormai mettendo a punto sistemi interattivi che usano la voce come mezzo di scambio di informazione in almeno una direzione, ed essi prolifereranno in un futuro abbastanza prossimo. Il loro successo dipende in qualche misura da una migliore capacità di sintesi e di riconoscimento del discorso, ma forse più specificatamente dal disegno di dialoghi che tengano conto delle caratteristiche del parlato e delle peculiari relazioni tra le persone e la parola parlata.

Si dispone già di un parlato artificiale di buona qualità, applicando il principio della concatenazione, ma la sua equivalenza a quello naturale dipende dalla disponibilità di memoria, per una riproduzione di buona qualità e per differenti versioni di unità vocali appropriate per i vari contesti. Il parlato sintetico è potenzialmente molto più flessibile ed economico, ma il suo utilizzo può aumentare i problemi umani d'assimilazione dell'informazione. Le caratteristiche della voce usata possono influenzare la visione che gli utenti hanno delle possibilità e della personalità del sistema.

La tecnologia del riconoscimento vocale continuerà e svilupparsi, dalle macchine addestrate a riconoscere e impiegare parole isolate e un vocabolario limitato, ai sistemi molto sofisticati, indipendenti dal parlatore e con vocabolari estesi, in grado di comprendere il linguaggio naturale. Questo obiettivo è ancora un po' lontano, dal momento che i problemi di riconoscimento vocale si stanno dimostrando assai più complessi di quelli di sintesi. Macchine di un livello intermedio di sofisticazione si possono dimostrare sino a un certo punto più difficili per l'utente di quelle molto semplici. Il progetto del dialogo è delicato ed ha bisogno di tenere conto delle limitazioni delle apparecchiature di riconoscimento, delle caratteristiche della capacità umana di elaborare informazioni e del comportamento conversazionale.

Pensare in parallelo

THEODOR HUETER (*)

SOMMARIO

Le architetture seriali che caratterizzano gli attuali computer e dispositivi di Input/Output possono essere state determinate, almeno in parte, da schemi culturali tipici del mondo occidentale.

Le culture orientali possono fornire lo spunto per progettare nuove interfacce uomo-macchina, organizzate in modo parallelo, adatte a realizzare concetti e obiettivi della Intelligenza Artificiale.

Considerate le mente umana come un computer biologico. Nelle culture occidentali l'input e output intellettuali tendono ad essere altamente seriali: le informazioni vengono trasmesse come una catena di fonemi e caratteri alfabetici.

Anche il nostro modo di pensare tende ad essere seriale; siamo infatti addestrati a sezionare un problema nei suoi elementi, e a costruire sequenze di relazioni causa-effetto tra tali elementi un passo alla volta.

In altre parole, sembriamo macchine di von Neumann organizzate serialmente. Ma questa organizzazione è in qualche modo preordinata ed inevitabile? Siamo cablati fin dalla nascita?

La testimonianza proveniente da varie fonti indica il contrario.

Le scritture a ideogrammi, come la scrittura cinese, fanno pensare che il computer biologico possa gestire input e output paralleli. E ci sono molti esempi di pensiero visivo anche nell'ingegneria e nella scienza occidentali.

Fondamentali scoperte scientifiche che implicano il pensiero non verbale o immagini mentali includono la scoperta dell'anello benzenico da parte di Auguste Kekule, la formulazione di DeBroglie del modello dell'atomo, e la visione della doppia elica del DNA da parte di Watson e Crick. Albert Einstein affermava di aver raramente pensato in parole, ma piuttosto in immagini visive.

In breve, l'evidenza fa pensare che noi abbiamo la capacità di elaborazione olistica di input multipli, una operazione altamente parallela.

Se ciò è vero, un esame più attento del computer biologico potrebbe fornire idee per nuovi approcci alla elaborazione parallela di informazioni, e per i nuovi tipi di interfacce uomo-macchina necessarie per l'intelligenza artificiale e per altre applicazioni.

Per ideare tali nuovi approcci potremmo dover abbandonare alcuni pregiudizi culturali. Nel mondo occidentale la logica sequenziale ha prevalso fin dai tempi di Aristotele, ed ha favorito storicamente soluzioni meccanicistiche a problemi scientifici.

Le culture orientali, al contrario, si sono basate soprattutto su processi visivi, olistici, non-logici. Le abitudini di linguaggio occidentali hanno alquanto compresso il pensiero visivo, favorendo il pensiero analitico o numerico.

I cinesi e i giapponesi, invece, impiegano simboli non-alfabetici che possono servire ugualmente bene (o meglio) per alcuni scopi.

Quanto profondamente radicata è la preferenza occidentale per i modi di pensiero seriali?

I modi verbali di comprensione della realtà dominarono in occidente durante tutto il Medioevo. Eugene S. Ferguson dell'Università del Delaware, il quale ha scritto la storia del pensiero non-verbale nella tecnologia, trova che i modi visivi di pensiero e di comunicazione di informazioni tecniche riguadagnarono impeto durante il Rinascimento, sebbene continuassero ad essere limitati dall'orientamento verbale e seriale del pensiero occidentale. [1].

Durante i secoli successivi, la corrente principale del pensiero scientifico divenne ancora meno grafica e concreta, spostandosi da ragionamenti geometrici (Descartes, Keplero) a ragionamenti aritmetici (Newton, Lagrange).

Alla fine del secolo scorso, William James osservò che la discussione filosofica si accentrava sulla domanda "è possibile il pensiero senza linguaggio?", domanda che poi dominò il dialogo tra Ludwing Wittgenstein e Bertrand Russell.

Nel 1977, tuttavia, Rudolph Arnheim, tra gli altri, aveva girato la do-

(*) Già V. President della Ricerca Centrale della Honeywell Inc.
L'articolo è tratto da *Scientific Honeyweller*, Vol. 7, n. 3, 1987.

manda e chiedeva: "Si può pensare in parole così come si può pensare in termini di cerchi, rettangoli, o altre forme simili? [2].

Quali sono le caratteristiche dei modi occidentali di comunicazione? La tabella di fig. 1 elenca diversi modi di comunicazione interpersonale, inclusi sia quelli verbali (parola e udito) che quelli visivi (scrittura e lettura).

Notiamo che le caratteristiche di funzionamento del canale acustico possono variare da banda stretta (parola) a banda larga (musica). Il canale visivo ha modi sia seriali (scansione) che paralleli (percezione dell'immagine).

Ovviamente non siamo riusciti a fornire ai nostri computer la stessa capacità di lavorare in parallelo che noi possiediamo. Nella lettura veloce, per esempio, le stringhe di parole sono trattate come immagini, ma i lettori di caratteri ottici trattano un carattere per volta. I nostri "algoritmi" di riconoscimento delle forme sono probabilmente altamente paralleli, ma nella maggior parte del riconoscimento computerizzato delle forme emerge che le immagini vengono sezionate in stringhe seriali di bit, un processo questo piuttosto innaturale.

La tabella di fig. 1 suggerisce, inoltre, che i modi occidentali di comunicazione non sfruttano le capacità parallele umane così compiutamente come invece fanno quelli orientali. Nelle culture occidentali, i modi di trasmissione risultano essere principalmente sequenziali, sebbene i modi di ricezione facciano un uso più completo delle intrinseche capacità di banda larga dei nostri sensi. Il sistema di codifica usato nel cinese scritto, d'altro lato, trae il massimo vantaggio dell'elaborazione parallela sia nei modi di trasmissione che in quelli di ricezione.

Fino a che punto i nostri modi sia di pensiero che di comunicazione sono condizionati dalla biologia? La neurofisiologia ha affrontato il problema di come il linguaggio sia gestito dai modi seriali e paralleli disponibili al cervello.

Gli studi di un certo numero di fenomeni psicologici, come la diversa dominanza degli emisferi cerebrali, fanno pensare che ci possa essere più variabilità di quanto si possa immaginare nell'organizzazione di queste funzioni e più flessibilità durante il loro sviluppo.

Il linguaggio, la concettualizzazione, la capacità visivo-spaziale e perfino il talento musicale, sono stati identificati come funzioni nelle quali la dominanza dell'uno o dell'altro emisfero cerebrale svolge un ruolo importante. La recente ricerca, che fa uso di radio-isotopi e di tomografia computerizzata, sta cominciando a quantificare le asimmetrie destra-sinistra nell'anatomia del cervello, e a correlare con modelli di comportamento, come l'essere destrorsi o mancini, l'abilità verbale, o i disordini di sviluppo nell'infanzia [3].

La parola, per esempio: si assume in generale che i "centri della parola" umani siano localizzati nella parte sinistra del cervello. Nonostante la caratteristica genetica di "controllo del linguaggio" da parte dell'emisfero sinistro del cervello, è stato anche provato che essa si instaura solo gradualmente durante l'infanzia (prima dei cinque anni) [4]. Si è anche scoperto che l'estensione di questa specializzazione emisferica e la misura della coordinazione tra emisferi provoca variazioni in abilità come la verbalizzazione e la risoluzione di problemi visivo-spaziali [5].

Gli studi neurologici e le ricerche di linguisti come Noam Chomsky per modellare processi cognitivi e linguistici, hanno stimolato recenti indagini sulle prime fasi di acquisizione del linguaggio da parte dei bambini [6]. Un tale lavoro sarebbe molto più producente se non fosse limitato a bambini con basi culturali occidentali [7]. Di particolare interesse sono i rapporti nella letteratura neuropsicologica che trattano la lettura e l'ascolto del giapponese [8]. La scrittura giapponese è una complessa mescolanza di ideogrammi *kanji* e di lettere fonetiche *kana*. Gli studi delle lesioni all'emisfero sinistro del cervello dimostrano che la capacità di leggere i kana segmentati serialmente viene persa prima della capacità di leggere i kanji ideografici [9]. Secondo Tsuneda ci sono inoltre differenze nel controllo cerebrale dei suoni del linguaggio. È stato scoperto che determinati stimoli musicali localizzati nell'emisfero destro del cervello in soggetti occidentali, interferiscono con le prestazioni linguistiche in soggetti orientali [10].

Si è portati a ritenere che l'integrazione delle funzioni coinvolte nel leggere, scrivere, ascoltare e parlare, dipenda in parte dalla cultura e dall'educazione linguistica.

Chiaramente c'è una differenza fondamentale tra le transazioni mentali richieste per comprendere gli ideogrammi e quelle richieste per comprendere i caratteri alfabetici; sebbene siano entrambi scritti come tratti sequenziali, l'elaborazione iniziale di un ideogramma è di tipo parallelo, mentre l'elaborazione iniziale di parole codificate alfabeticamente è seriale. La lettura veloce, ossia la capacità di leggere parole o frasi complete senza scandirle, è attuata solo da una piccola percentuale di occidentali.

Studi su bambini americani affetti da dislessia, un disordine nella lettura, hanno rivelato una capacità notevolmente maggiore di leggere pittografie piuttosto che parole alfabetiche [11]. Studi comparativi simili in Giappone indicano che i kanji sono elaborati soprattutto nell'emisfero destro del cervello, mentre i kana, come la parola, sono gestiti dall'emisfero sinistro.

Queste osservazioni inducono a concludere che gli ideogrammi come i kanji si accordano bene col riconoscimento delle forme e sono caratterizzati da processi di estrazione del sistema retino-corticale che sono regolati per affinare la discriminazione e l'orientamento delle linee [12].

Tali osservazioni possono inoltre dare un certo peso alla congettura che, sebbene questi processi di base si sviluppino neurogeneticamente, i processi di più alto livello per l'assemblaggio e la correlazione di modelli visivi, siano affinati dalle esperienze di apprendimento durante la prima infanzia, e perciò soggetti a condizionamenti culturali [13].

La capacità di vedere intere frasi alla volta implica che sia nella scrittura

Fig. 1 - I modi di comunicazione umana evidenziano la nostra capacità sia di trasmettere che di ricevere le informazioni in una forma compressa, o parallela, e come le culture orientali usino questa capacità in misura maggiore di quelle occidentali.

		SEQUENZIALE	PARALLELO
TRASMISSIONE	MECCANICA	STAMPA DELL'INGLESE SCRITTURA MANUALE	STAMPA DEL CINESE
	ACUSTICA	SOPRATTUTTO LA PAROLA	LUNGUAGGI TONALI MUSICA
	OTTICA	SEMAFORO	
RICEZIONE	ACUSTICA	SOPRATTUTTO L'ASCOLTO	PERCEZIONE DELLA MUSICA
	OTTICA	LETTURA DI SCRITTI INGLESI SCANSIONE	LETTURA DI SCRITTI CINESI LETTURA VELOCE PERCEZIONE DI IMMAGINI
	TATTILE		BRAILLE

Fig. 2 - Il conteggio dei tratti per una riga di versi dà un'idea della compressione dei dati resa possibile dagli ideogrammi cinesi.

IDEOGRAMMA	TRATTI	EQUIVALENTE INGLESE	TRATTI
毋忘在莒	6	DO NOT FORGET	20
	9	NATIONAL HUMILIATION (DURING)	37
	6	PEACE (AND)	17
	12	SECURITY	15
	33		89

che nella lettura di diagrammi possa aver luogo un certo tipo di *compressione di dati*. È stato suggerito all'autore che la poesia classica cinese può dare qualche idea al riguardo [14]. Contando il numero di tratti richiesto per la trascrizione sia cinese che inglese di una poesia della dinastia T'ang, si può ottenere una stima del fattore di compressione-dati della scrittura [15]. Naturalmente, ci può essere molta più compressione di dati durante la lettura che durante la scrittura, in quanto la scrittura è un processo intrinsecamente seriale.
La tabella di fig. 2 confronta il numero di tratti richiesto per trascrivere una riga di versi in cinese e in inglese. In questo caso il rapporto di compressione è tre.
Diverse poesie di varia lunghezza e contenuto, scelte a caso, presentavano rapporti tra due e tre. Naturalmente, anche la sintassi e la semantica, che differiscono molto tra verso e prosa, possono influenzare questi rapporti.

Anche la compressione di dati ha luogo nel cervello? Qui i messaggi sono decodificati e generati con elaborazione associativa, un'operazione altamente parallela resa possibile dalla straordinaria connettività del cervello [16]. Dovrebbe essere possibile ideare test che misurino il tempo di transazione richiesto per la lettura veloce alfabetica e per quella ideografica. I risultati dipenderebbero probabilmente dalla sintassi e dalla semantica dei messaggi di test. Ma il tempo di transazione potrebbe anche rivelarsi dipendente dai fattori culturali, poichè le necessità di bufferizzazione e di correlazione connesse alle conversioni da seriale a parallelo possono essere soddisfatte da matrici di connettività che sono condizionate culturalmente [17].

Se una tale dipendenza culturale può venire dimostrata, sarebbe necessaria una ulteriore ricerca per delimitare il suo ambito e per determinare dove nuovi simboli non alfabetici, o sistemi iconici, possono essere utili per la comunicazione uomo-macchina.

Si suggerisce di eseguire studi durante i quali gruppi selezionati di studenti vengono esposti a differenti tipi di informazioni codificate sia ideograficamente che alfabeticamente.

Si potrebbero cercare le differenze nel tempo richiesto per l'immagazzinamento e il recupero delle informazioni, osservare la completezza con cui le informazioni possono essere immagazzinate e recuperate e le differenze tra questi parametri in studenti con differenti basi culturali.

La capacità dei giapponesi di gestire due diversi tipi di codici costituisce un'opportunità singolare. Non solo i kana sembrano essere più soggetti

alla dominanza linguistica della parte sinistra del cervello, ma la moderna neurologia avvalora anche l'ipotesi che due tipi di caratteri possono essere elaborati da aree differenti nell'emisfero sinistro. L'area di Broca (fig. 3), che controlla la vocalizzazione, può essere più importante nell'elaborazione dei kana, e l'area di Wernicke, che facilita la comprensione di parole scritte e parlate e la formattazione delle espressioni corrispondenti - attività presumibilmente altamente parallele - può essere più importante nell'elaborazione dei kanji.

Entrambe le aree sono coordinate dall'ippocampo, il quale controlla il trasferimento da memoria a breve-termine a memoria a lungo-termine [18].

La recente crescita esplosiva dei modi di input e output grafici per i computer addita i vantaggi potenziali di queste ricerche.

Se il modo visivo è davvero il modo primario e più spontaneo per la comprensione umana, la limitazione a modi seriali della maggior parte degli input e output di computer può essere alquanto artificiale.

Per sfruttare pienamente la nostra capacità di elaborazione parallela, dobbiamo cercare di estendere la larghezza di banda dei canali attraverso i quali comunichiamo con le macchine. Ciò viene ora tentato in una vasta gamma di aree, dai giochi elettronici al progetto con l'aiuto del computer. Ma il modo in cui trasmettiamo istruzioni o comandi ad una macchina lascia ancora molto a desiderare. Perfino un più largo uso di linguaggi ad alto livello, come il Lisp o il Prolog, con il loro software continuamente in sviluppo, può non essere la migliore risposta, una volta sviluppati codici visivi, o ideografici, adeguati.

Gli input di dati provenienti da organi sensori, come i lettori di caratteri ottici, sono anch'essi sequenziali, e una manipolazione dei dati più parallela, come nelle macchine MIMD (Multiple Instruction/Multiple Data stream), è ancora oggetto di ricerca. La presentazione di dati sotto forma di codici visivi di tipo iconico o facendo uso del colore si è diffusa largamente, ma richiede ancora un migliore adattamento alle capacità umane. La tabella di fig. 4 evidenzia come le nuove tecnologie, quali l'elaborazione ottica e la memorizzazione olografica di informazioni, possano aiutare a superare queste limitazioni.

È opinione di chi scrive che la comprensione del funzionamento dei modi di linguaggio seriale e parallelo in culture differenti può far luce su come i computer e gli esseri umani possano interagire al meglio.

Da numerosi indizi risulta, inoltre, che la ricerca dedicata ai modi paralleli di elaborazione del linguaggio potrebbe essere ricca di risultati interessanti, sia per la neuropsicologia che per la scienza informatica.

Può essere certamente vero che le macchine di von Neumann organizzate serialmente abbiano limitazioni intrinseche imposte dalle inclinazioni culturali e dalle predisposizioni neuro-architetturali dei loro inventori e gestori.

Nonostante il notevole successo di queste macchine "occidentali" durante gli ultimi decenni, possiamo raggiungere altri importanti traguardi se tentiamo di ideare una architettura di computer più cosmopolita. La constatazione e la comprensione dei nostri limiti culturali possono condurre ad inaspettati sviluppi futuri.

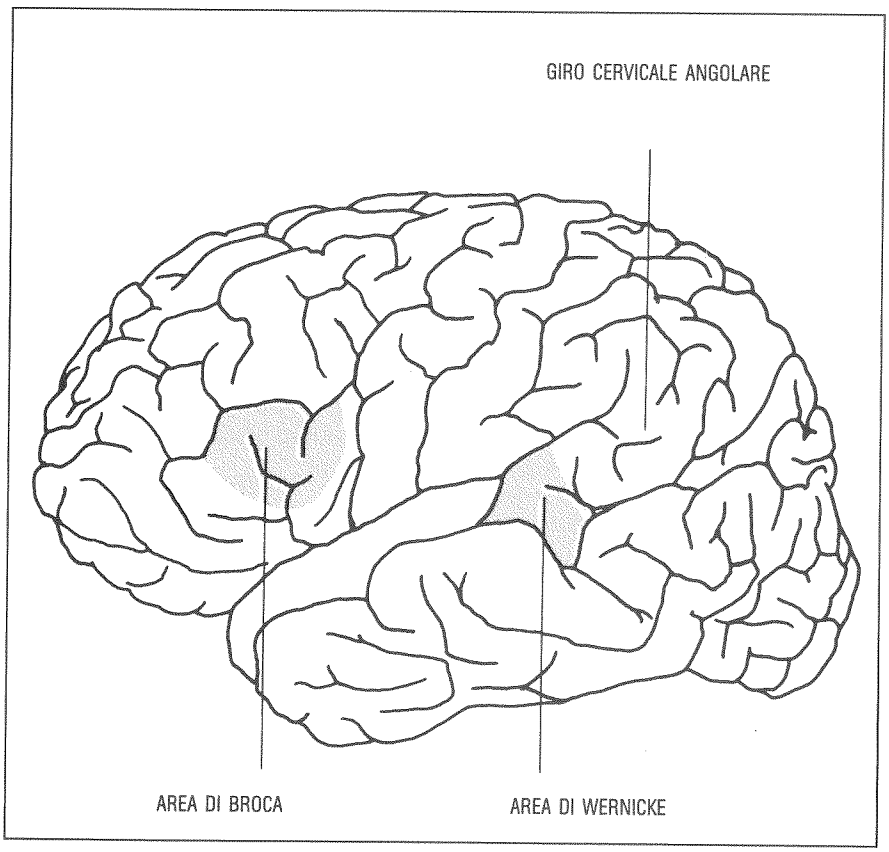


Fig. 3 - I centri della parola nell'emisfero sinistro del cervello includono l'area di Broca e l'area di Wernicke - che sono coinvolte nella comprensione e produzione della parola - e il giro angolare - che media tra informazioni scritte e parlate.
È interessante notare che le aree suddette sembrano avere una diversa importanza nell'elaborazione dei kanji - gli ideogrammi usati per scrivere radici di parole in giapponese - e dei kana - le lettere fonetiche usate principalmente per finali di parole. L'area di Werniche sembra, infatti, svolgere un ruolo più importante nell'elaborazione dei kanji, e l'area di Broca nell'elaborazione dei kana. Probabilmente la scrittura ideografica permette al cervello di fare migliore uso della sua capacità di elaborazione parallela.

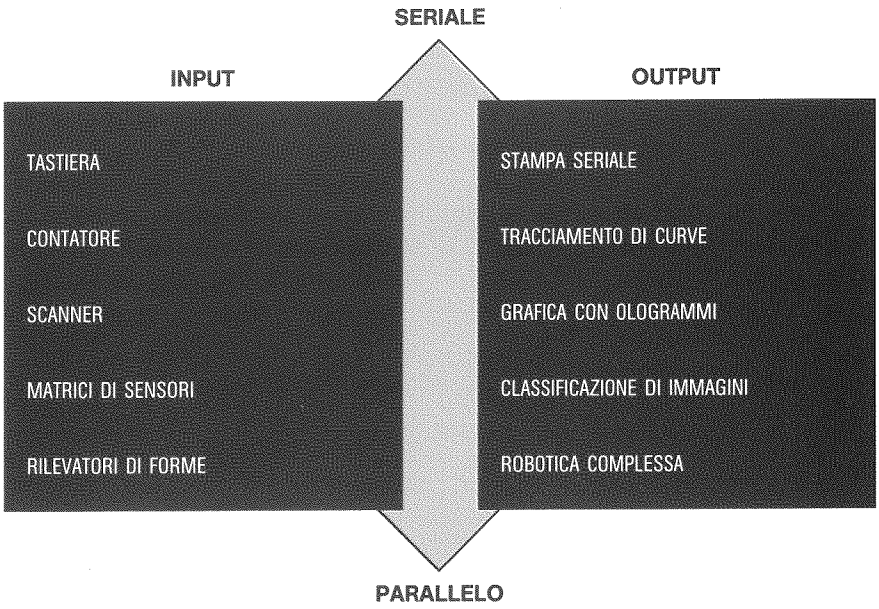


Fig. 4 - I modi di input/output dei computer sono caratterizzati da un elevato grado di serialità. Le nuove tecnologie consentono un maggiore parallelismo, ma si impongono ulteriori progressi.

Ringraziamenti

L'autore esprime la sua gratitudine a Chente Chao, che ha fatto parte del Centro di Scienze Fisiche della Honeywell, per la sue ricerche sugli ideogrammi cinesi, a William Charlsworth, dell'Università del Minnesota, per gli studi dei modi linguistici, e a Chris Reiner per la revisione critica della prima stesura di questo articolo.

Bibliografia

[1] E.S. FERGUSON, "The Mind's Eye: Nonverbal Thought in Technology", *Science,* Vol. 197, 1977, p. 827.

[2] R. ARNHEIM, *Visual Thinking,* Faber and Faber Ltd., London, 1977, p. 227.

[3] A.M. GALABURDA et al., "Right-Left Symmetries in the Brain", *Science,* Vol. 199, 1978, P. 852.

[4] L.S. BASSER, "Hemiplegia of Early Onset and the Faculty of Speech", *Brain,* Vol. 85, 1962, p. 427.

[5] According to R. Arnheim (loc. cit.) "language stabilizes perceptual imagery". Coordination between hemispheres may by the mechanism by which this is accomplished.

[6] N. CHOMSKY, *Rules and Representations,* Columbia University Press, New York, 1980.

[7] T. SAKAMOTO, M. MAKITA, "Cross-National Studies of Behavior and Processes in Reading and Writing", *Comparative Reading* (John Downing, editor), Macmillan, New York, 1973.

[8] H. GOYO, T. ADACHI et. al., "Late Positive Component during Semantic Information Processing in Kanji and Kana Words", *Evoked Potential,* (Colin Baker, editor), University Park Press, Baltimore, 1980.

[9] T. HATTA, "Recognition of Japanese Kanji in the Left and Right Visual Fields, *"Neurospychologica.,* Vol. 15, 1977, p. 687.

[15] P. NAITON, "The Japanese Brain: The Tsuneda Method", *ONR Far East Scientific Bulletin,* Vol. 9, No. 1, 1984, p. 111.

[11] P. ROZIN, S. PORTISKY and R. SOTSKY, "American Children With Reading Problems Can Easily Learn to Read English Represented by Chinese Caracters, *Science,* Vol. 171, 1971, p. 1264.

[12] D. H. HUBEL and T. N. WIESEL, "Receptive Fields in the Cat's Visual Cortex", *Journal of Physiology,* Vol. 160, 1962, p. 106.

[13] E.J. GIBSON and H. LEVIN, *The Psychology of Reading,* M.I.T. Press, 1975, p. 111.

[14] GODWING C. CHU., The East-West Center, Honolulu, Hawaii, private communication.

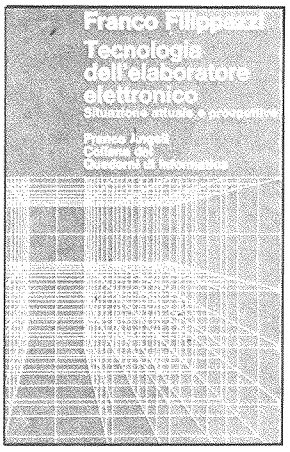
[15] W. BYNNER, *Three Hundred Poems of the T'ang Dynasty,* Book World Company, Taipei, Republic of China, 1963.

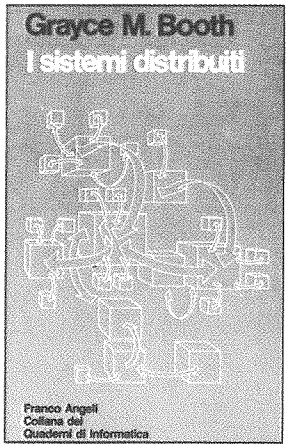
[16] N. GESCHWIND, "Language and the Brain", *Scientific American,* Vol. 226, 1972, p. 178.

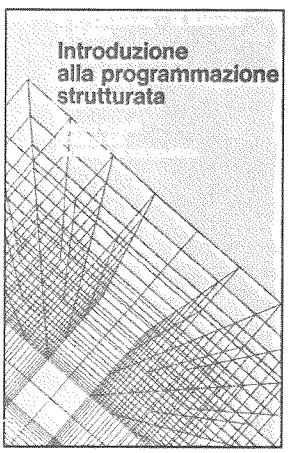
[17] C.J. LUMSDEN and E. O. WILSON, *Genes, Mind, Culture,* Harvard University Press, 1981.

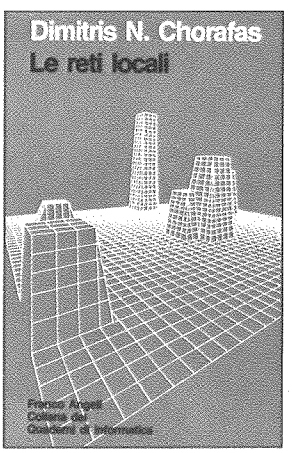
[18] J. O'KEEFE and L. NADEL, *The Hippocampus as a Cognitive Map,* Clarendon Press, New York, 1979.

I TESTI DEI
QUADERNI DI INFORMATICA

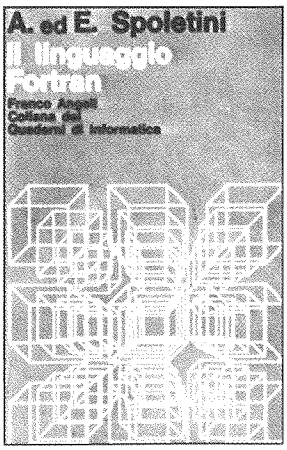


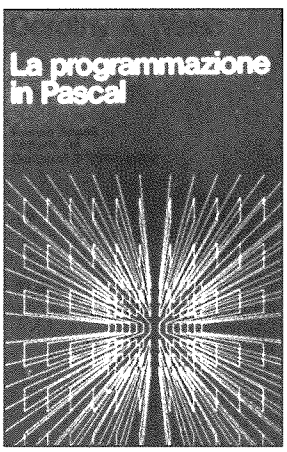


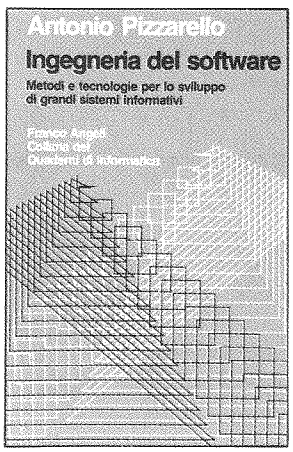


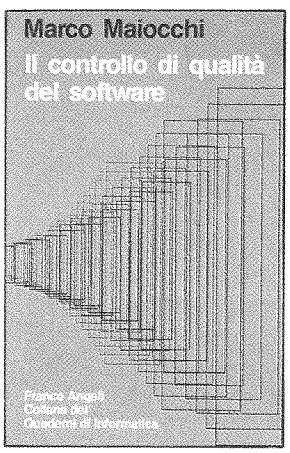


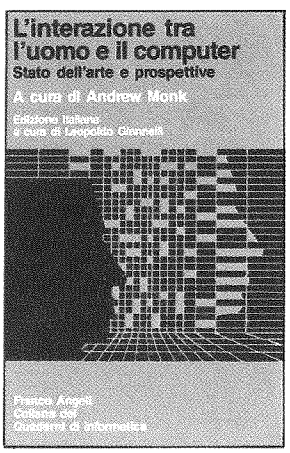
I TESTI DEI
QUADERNI DI INFORMATICA











Spedizione in abbonamento postale a tariffa intera/Milano